POSTER ABSTRACTS

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

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

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

POSTER AUTHOR SCHEDULE

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

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

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

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

Objectives: We aimed to develop methodology for epigenome-wide association studies (EWAS) of DNA methylation. This includes assessments and optimisation of current approaches for data preprocessing and on large-scale datasets (N=2,000). Methods: The EpiMigrant DNA methylation data set was generated using the Illumina HM450 platform and includes baseline DNA samples of non-diabetic South Asians (N=2,687), approximately half of whom went on to develop diabetes. Of these samples, 36 were measured in duplicate to constitute the replication subset. We filtered methylation scores by a Bonferroni-corrected detection p-value, cutoff, instead of the default value of p<0.05. As a metric of performance, we used the count of false methylation calls on the Y chromosome in females. To test whether results were valid for autosomal markers, we compared the autosomal per-marker call rate to the rate of outliers (beta > 1.5 inter-quartile range). We then assessed two different methods for normalisation of methylation scores across arrays using the probe-wise correlation between duplicates in the technical replication dataset. We used three normalisation methods: QN1, quantile normalisation on intensity values, further separated by methylated or unmethylated targets into six categories. Results: We find that using the Bonferroni correction threshold (p<10e-6) and using the count of false methylation calls on the Y chromosome in females. Remaining intensity signals can be explained by probes cross-hybridizing to autosomal regions. For autosomal markers, using a lower detection p-value cutoff decreases the outlier rate while only filtering 0.4% of markers. We find that before normalisation, only 9.9% of markers show high correlation (r>0.8) between the two duplicates. QN1 increases this proportion to 10.6%, while QN 2 leads to 21.8% of probes reproducing (r>0.8). The remaining high percentage of weakly performing probes can be explained by experimental variation exceeding inter-individual variation in methylation. These invariant markers are also unlikely to give rise to an association signal. Conclusion: We conclude that any analysis of large-scale EWAS data should be based on probes filtered for a Bonferroni-corrected detection p-value. To further increase data quality, we recommend separate quantile normalisation on intensities of the six different probe categories.

414T  Epigenome-wide DNA methylation study reveals hypermethylated colla gen genes and suggests a role for TGFβ in osteoarthritis. M.A. Jef fress1,*, J.A. James2, H. Sawalha1, 1) Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Arthritis & Immunology, OMRF, Oklahoma City, OK; 3) Rheumatology, University of Michigan, Ann Arbor, MI.

Background: Osteoarthritis (OA) is the leading cause of chronic disability in the U.S., affecting 40% of individuals over the age of 70 and costing $128 billion annually in the US alone. Late-stage OA chondrocytes exhibit a host of gene transcription changes leading to upregulation of enzymes that contribute to cartilage breakdown. Herein, we characterize epigenome-wide DNA methylation changes in osteoarthritis compared to healthy cartilage from the same joints. Methods: Articular cartilage tissue from 12 OA femoral heads was dissected from affected and normal areas, frozen in liquid nitrogen, and DNA extracted. Following sodium bisulfite-treatment, DNA methyla- tion was quantified at >485,000 CpG sites across the genome using Illumina HumanMethylation450 arrays. CpG sites with an absolute methylation differ- ence between OA and normal cartilage (Δβ) > 15%, and P < 0.01 after correction for multiple testing using a false discovery rate of 5%, were considered significant and used for analysis. Results: We found 442 differentially methylated CpG sites; 260 hypo- and 182 hypermethylated. Overrepre- sented gene sets included ‘Connective tissue disorder’ (n=53, p=4.73E-6 to 7.36E-3), ‘Developmental disorder’ (n=44, p=4.73E-6 to 7.36E-3), and ‘Skeletal & muscular disorder’ (n=59, p=4.73E-6 to 7.36E-3). Interesting DM in OA include hypermethylation of COL11A2, which functions to maintain spacing and diameter of type II collagen and is mutated in patients with Stickler syndrome and OSMED. Additionally, we found hypermethylation of COL2A2, hypermethylation of the fibrillar collagen gene COL1A1, and hypermethylation of COL18A1. Hypermethylation was also noted within the WNT pathway coreceptor LRP5, associated with OA in mice and is a suscepti- bility gene for OA in humans. Hypomethylation was found in the transcrip- tion factor RUNX1. Upstream regulator analysis identified significant associ- ation of TGFβ1 with 44 differentially methylated genes. Finally, canonical pathway analysis identified enrichment of several pathways, most promi- nently the ERK signaling pathway among differentially methylated genes (n=27, p=4.41E-6 to 6.93E-4). Overall, our hypothesis using an epigenome-wide association approach. Methods: We used a total of 17 patients having knee- or hip-joint replacement due to primary OA and 24 healthy controls participated in this study. DNA was extracted from blood. DNA methylation profiling was performed using the Illumina Infinium HumanMethylation450 chip, which measures about 480,000 differ- ent CpG sites covering 96% of RefSeq genes. It provides comprehensive gene region coverage, targeting multiple sites including the promoter, 5′ UTR, 1st exon, gene body and 3′ UTR. The methylation level at each CpG site was measured by β values varying from 0 (no methylation) to 1 (100% methylation). The significance level was defined as p < 1.3 x 10−7 after Bonferroni correction for multiple testing. Results: We found 539 individual CpG sites significantly associated with OA, with a mean difference in methylation level of >10%. Among them, 516 were hypermethylated and 23 were hypomethylated in OA patients. These individual CpG sites represent 322 independent gene regions. Although methylation patterns may be tissue specific, 15 of the CpG regions that we identified were recently found to also be differentially methylated in articular chondrocytes from OA patients (Ann Rheum Dis 2013 Mar 16). Conclusion: We demonstrate that DNA methylation patterns in blood are associated with healthy and OA healthy controls. This may be valuable for identifying those at high risk for OA and the development of effective treatment for OA.
416T

Diurnal Rhythms of Clock Gene DNA Methylation and their Relationship to Rhythms of Clock Gene Expression in the Human Cerebral Cortex. A.S. Lim1, G.P. Schenkel1, L. Yu1, B. Buchman2, J.A. Schneider3, A.J. Myers4, D.A. Bennett5, P.L. De Jager6, 1) Division of Neurology, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada; 2) Department of Neurology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 4) Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago, IL; 5) Department of Psychiatry, University of Miami, Miami, FL.

SACKRO2 et al. demonstrated that circadian clock is regulated by a highly conserved series of ‘clock’ genes participating in a near 24-h transcription-translation-negative feedback loop. Recent work in Neurospora has suggested that rhythms of DNA methylation in or near clock genes may play a role in circadian clock regulation. However, whether such rhythms are also present in human tissues, and how they relate to clock gene expression, is uncertain. METHODS: We quantified DNA methylation at 128 CpG sites in or near 6 canonical clock genes - PER2, PER3, CRY1, CRY2, ARNTL, and CLOCK - using Illumina Infinium HumanMethylation450 microarray data from dorsolateral prefrontal cortex samples from 735 deceased individuals in 2 cohort studies of older individuals, the Religious Orders Study and the Rush Memory and Aging Project. We quantified transcript abundance for these genes using Illumina Human HT-12 Expression Microarray data from a subset of 490 of these individuals. Transcript abundance and methylation level at each CpG site was parameterized as a function of time of death using cosine curves. RESULTS: Significant daily rhythms of methylation were seen in 63/128 CpG sites (p<0.05), and in 3/6 genes. PER2 (p=1.9×10−4), PER3 (p=1.0×10−4), CRY1 (p=7.0×10−4), ARNTL (p=5.9×10−5), and CLOCK (p=7.3×10−3) with the timings of peak transcript abundance mirroring those seen in other diurnal mammals. The timing of the nadir of methylation was specific. For rhythmic Clock genes, the timing of the nadir of methylation clustered between 16:00 and 22:00, irrespective of the timing of transcript abundance. However, for sites upstream of the promoter region and in the 5’UTR, the timing of the nadir of methylation was roughly in-phase with the timing of peak abundance of the corresponding transcript. This suggests temporal correlation between hypomethylation and expression. Finally, significant sex differences were seen, with transcript abundance peaking later in men than in women for all 5 rhythmic clock genes, and the nadir of methylation occurring later in men than in women.

CpG sites. CONCLUSIONS: There are site-specific diurnal rhythms of DNA methylation in and near canonical clock genes, with characteristic site-specific phase relationships to transcript abundance and to global circadian rhythms. Rhythms of DNA methylation may play a role in the regulation of the circadian clock in human cerebral cortex.

417F

DNA methylation analysis of human neuronal and non-neuronal cells at the base pair resolution. J. Ueda1, M. Lundmark2, S. Kashiwaba3, F. Ikeda3, M. Ichikawa4, J. Ishii5, Y. Yokokawa4, T. Iwamoto4, M. Shinozaki4, T. Kato1, 1) Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, Wako, Japan; 2) Department of Molecular Psychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 4) Department of Neuropsychiatry, School of Medicine, Sapporo Medical University, Sapporo, JAPAN.

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

418T

Methylation of Leukocyte DNA and Ovarian Cancer: Relationship with Disease Status and Outcome. B. Fridley1, S. Armasu2, M. Ciccek2, K. Kalay2, M. Larson3, D. Koessler4, D. Rider5, V. Shridhar2, J. Olson3, J. Cunningham5, E. Goode2. 1) University of Kansas Medical Center, Kansas City, KS, USA; 2) Mayo Clinic, Rochester, MN, USA; 3) Dartmouth College, Lebanon, NH, USA.

We hypothesized that DNA methylation (DNAm) in leukocytes of epithelial ovarian cancer (EOC) cases and controls may differ and that, among cases, this DNAm may vary by disease outcome. Thus, we performed case-control and survival analyses using 336 EOC cases and 398 controls (3 experimental batches) with blood-based DNAm assayed with the Illumina Infinium HumanMethylation27 and 450 BeadChips. To limit spurious results, we limited analysis to CpG probes measured on both arrays, and we removed non-specific probes and probes associated with distributions of white blood cell types. The association of the DNAm levels for 24,520 CpG sites with disease status and overall survival (OS) was completed using linear and Cox-proportional hazards models. Meta-analysis across three batches was completed using a random effect meta-analysis. The top association between DNAm and case-control status was observed for CpG cg04834572 near DUSP13 (p=1.6×10−14, individual batch p range from 2.1×10−4 to 1.1×10−6). The top pathway enriched, based on Fisher’s Exact test, was the telomerase signaling (p=1.24×10−3). This pathway and variants in TERT have been found to be associated with the development of many cancers, including EOC. TERT has an extensive role in the maintenance of functional telomeres which are important in the protection of chromosomes from DNA damage. The top CpG sites associated with OS were: cg10276549 near GABRE (p = 5.8×10−5); cg06171242 near ACOT13 and TTRAP (p=4.4×10−4); and cg06171247 near ACOT13, TTRAP, and PPIG (p=1.5×10−4). The most enriched pathways were: relaxin signaling (p=7.09×10−5) and CXC4 signal (p=1.25×10−4) and IL-8 signaling (p=3.05×10−4), all containing GNB1, GNA12, and PIK3R4. GNA12 has been found to be associated with response to platinum/taxane combination therapy, while PIK3R4 is a member of the phosphoinositide-3-kinases family which is involved in multiple cell functions, including proliferation and cell survival. Via a CpG-by-Cpg approach accounting for CpGs known to correlate with white blood cell types, we identified CpG sites that point to potential DNAm for EOC status and OS. Although pre-treatment case samples were used, a study limitation is the lack of prediagnostic blood samples for EOC. Thus, we are unable to conclude that the CpGs are involved in disease etiology. Future research is needed to functionally determine the biological and epigenetic relevance of the detected CpG sites associated with EOC status and OS.

419F

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

The association of genetic abnormalities and specific cancers is leading to an increasing number of bio-markers that can be used for early detection of cancers. A number of studies have been performed to determine some biomarkers and to establish diagnostic criteria. Additionally, the methylation state of an increasing number of genes is also being produced possibly predictive bio-markers for specific cancers and stages of cancer as well. To this end, we have analyzed both breast and ovarian cancers and their normal adjacent tissue using both DNA methylation analysis and genome wide array screening for LOH and copy number variations. The combination of genomic and epigenomic analysis in the determination of treatment protocols and predicted prognoses. Via a CpG-by-Cpg approach we have also identified genomic regions that may be producing possible predictive bio-markers for specific cancers and stages of cancer as well. To this end, we have analyzed both breast and ovarian cancers and their normal adjacent tissue using both DNA methylation analysis and genome wide array screening for LOH and copy number variations. The combination of these two methods has yielded several groupings of markers that are cancer specific as well as showing specificity for the specific tumor type. We will present the results of the combination of the genome wide analysis and DNA methylation results that may be used in combination to produce potential biomarkers for these two cancer types.
Independent contribution of epigenetic modifications within lipoprotein metabolism genes to plasma lipid profile variability. S.P. Guay 1,2, D. Brisson 1, B. Lamarche 2, D. Gaudet 2, L. Bouchard 1, 1) Department of Biochemistry, Université de Sherbrooke, Sherbrooke, QC, Canada; 2) ECOC GENE-21 and Lipid Clinic, Chicoutimi Hospital, Saguenay, QC, Canada; 3) Department of Medicine, Université de Montréal, Montréal, QC, Canada; 4) Institute of Nutrition and Functional Foods, Université Laval, Québec, QC, Canada.

Background: Inheritance plays a central role in the determination of plasma levels of lipids by explaining up to 60% of the interindividual variability. However, the gene polymorphisms identified so far explain less than 25% of the heritability of plasma lipid levels. Recent studies suggest that epigenetic modifications (DNA methylation), a non-traditional hereditary factor, could explain a significant proportion of the missing heritability of complex traits, such as plasma lipid levels. Objective: To assess whether the DNA methylation of key genes of the lipoprotein metabolism is associated with changes in fasting plasma lipid levels (high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and triglycerides (TG)) in patients with familial hypercholesterolemia (FH) carrying the same LDLR mutation (p.W68R).

Methods/Results: In the current study, 98 untreated FH patients (61 men and 37 women) were recruited. Blood DNA methylation levels were measured at the ABCA1, ABCG1, CETP, LCAT, LDLR, LIPC, LDLR, PLTP and SCARB1 gene loci using bisulfite pyrosequencing. Partial Pearson’s correlation analysis showed that DNA methylation levels at the ABCA1, ABCG1, CETP, LIPC, LDLR and PLTP gene loci were significantly associated with HDL-C, LDL-C and/or TG levels in a sex-specific manner in FH (all p<0.05). Multivariate linear regression models showed that conventional predictors of the plasma lipid profile (age, anthropometric measurements, blood pressure and fasting glucose levels) explained 33.7% of HDL-C, 17.5% of LDL-C and 26.1% of TG level variability in men (all p<0.001), and 11.4% of HDL-C, 10.7% of LDL-C and 25.3% of TG level variability in women (all p<0.02). When epivariants were added to the statistical models, the variance explained by the combination of conventional and epigenetic factors was found to be increased for both sexes (49.9% for HDL-C, 38.2% for LDL-C and 33.0% for TG levels in men (all p<0.001), and 16.8% for HDL-C, 14.9% for LDL-C and 41.8% for TG levels in women (all p<0.001). Together, these epivariants independently explained up to 15.8% of HDL-C, 24.0% of LDL-C and 14.7% of TG level variability in FH (all p<0.02).

Conclusion: These results suggest that epigenetic perturbations of key lipoprotein metabolism genes are associated with fasting HDL-C, LDL-C and/or TG levels in a sex-specific manner.
424T
Dose-dependent effect of in utero smoking on DNA methylation among Latino children in a methylation-wide association study. S.S. Oh,1 D. Hu1, C.P. Gignoux2, J.M. Galanter1, S. Huntsman3, D. Torgerson1, C. Eng3, L.A. Roth1, Á. Davis3, H.J. Farber3, P.C. Avila2, E. Brighina-Buenaventura2, M.A. LeNoir4, K. Meade2, D. Serekbrsky5, L.N. Borrell6, W. Rodriguez-Cintron1, R. Kumar1, J.R. Rodriguez-Santana1, F. Lurmann1, E. Burnham1, 1) UC San Francisco, San Francisco, CA; 2) Children’s Hospital and Research Center Oakland, Oakland, CA; 3) Baylor College of Medicine and Texas Children’s Hospital, Houston, TX; 4) Feinberg School of Medicine, Northwestern University, Chicago, IL; 5) Kaiser Permanente-Vallejo Medical Center, Vallejo, CA; 6) Bay Area Pediatrics, Oakland, CA; 7) Jacobi Medical Center, Bronx, NY; 8) City University of New York, Bronx, NY; 9) Veterans Caribbean Health System, San Juan, Puerto Rico; 10) The Ann and Robert H. Lurie Children’s Hospital of Chicago, Chicago, IL; 11) Centro de Neurologia Pediatrica, San Juan, Puerto Rico; 12) Sono Technology, Petaluma, CA.

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

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

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

425F

Type 2 diabetes (T2D) is a global epidemic, becoming increasingly prevalent due to a rising incidence of obesity caused primarily by poor diet and lack of physical activity. The incidence of T2D is particularly high in Mexican Americans and although the disease is known to be genetically regulated, implicated loci explain only a small portion of the genetic liability. Epigenetic regulation, such as DNA methylation, is a novel mechanism that may lead to gene dysfunction and disease development. Using Illumina HumanMethylation450 BeadChips, we performed genome-wide DNA methylation profiling of >450,000 CpG sites in peripheral blood cells from 859 Mexican Americans from ~40 large pedigrees. In total, 20% of individuals were diagnosed with T2D and all individuals had fasting glucose measurements available. Normalized DNA methylation data underwent analysis using SOLAR to test for heritability of each CpG site, and for association with T2D status and fasting glucose (in non-diabetic samples only). We used the Bonferroni method to correct for multiple comparisons. Approximately 24% of CpG sites were found to be significantly heritable (mean h²=0.453, p<1.06×10⁻⁷). We identified 12 statistically significant associations between DNA methylation levels and T2D (p<1.04×10⁻⁷). The top association was for a CpG site in the TXNIP gene (p=5.57×10⁻¹⁰), TXNIP is known to modulate glucose metabolism and insulin sensitivity, likely playing a role in T2D and possibly contributing to the therapeutic action of metformin. We also identified a significant association between diabetes and ABCG1 (p=5.45×10⁻¹⁰), which has been implicated in reverse cholesterol transport, insulin secretion and glucose tolerance. Significant associations between DNA methylation and T2D status were also seen in other genes previously implicated in diabetes, including CPT1A (p=5.92×10⁻¹⁰) and DHCR24 (p=7.66×10⁻⁹), as well as in potentially novel genes. Most of these genes were also at least nominally significantly associated with fasting glucose in the non-diabetic portion of our dataset (n=683). Our study has identified a number of genes that show DNA methylation changes associated with T2D in a Mexican American population. We are currently conducting pyrosequencing studies to validate these findings. We hypothesize that DNA methylation changes associated with T2D may identify a new set of therapeutic targets to delay the onset of T2D through lifestyle changes.

426T
Methylation analysis of CpG islands in the TYR and P genes: correlation with gene expression. X. Wang1, L. DSOUDA1, H. HE1, C. ANTOLIK1, Q. ZHANG1, 1. OGVFB, NEI/NIH, Bethesda, MD; 2. Department of Immunology, Capital Medical University, Beijing, China.

Oculocutaneous albinism types 1 and 2 (OCA1 & OCA2) are autosomal recessive disorders of melanin biosynthesis characterized by hypopigmentation of hair, skin and eyes. OCA1 is caused by mutations in the gene that encodes tyrosinase (TYR; 11q14-q21). OCA2 is caused by mutations in the gene that encodes the P protein (OCA2; 15q11.2-q12). Mutations in these genes have been detected in a majority of OCA patients, but a small proportion do not have any detectable mutations or have only one identifiable mutation despite sequencing all the coding regions of the gene. Epigenetic phenomena have demonstrated roles in the regulation of gene expression. However, the role of DNA methylation in the development of OCA is not known. Three melanoma cell lines (A375, SK-28, and MNT1) and a retinal pigment epithelium cell line (ARPE-19) were previously found to contain different levels of pigmentation and gene expression. The A375 cell line has no TYR and P gene expression and pigmentation. The SK-28 cell line has no TYR and P gene expression and pigmentation. The MNT1 cell line has no TYR and P gene expression but has pigmentation. The ARPE-19 cell line has pigmentation and TYR gene expression but no P gene expression. We hypothesized that the DNA methylation status in the candidate CpG islands around promoter regions of the TYR and P genes may play a role in the regulation of gene expression. Therefore, we analyzed TYR and P gene CpG methylation status by bisulfite sequencing of genomic DNA from these cell lines. We found CpG sites to be hypermethylated in TYR in the A375, SK-28 and ARPE-19 cell lines. In contrast, those CpG sites were hypomethylated in the MNT1 cell line. Further, we found that only the A375 cell line was hypermethylated (all the CpG sites were methylated for the P gene CpG island, and the rest of the cell lines had a hypomethylated CpG island (all the CG sites were unmethylated). Corresponding mRNA levels were also correlated by qRT-PCR. Further analysis should lead to a better understanding of the role of methylation in the regulation of TYR and P gene expression.
427F The Role of Brain DNA Methylation in the Pathology of Alzheimer’s Disease: Evidence of an Interaction Effect. PL: De Jager1,2,3, G. Srivas-tava1,2,4, W. Eaton1,2, M. Yang1, L. Yu1, S. Kuiper2, Y. Palma-ciova3, M. Kellis2,4, DA. Bennett5, LB. Chibnik2,3, J. Kinne1,2,3) 1) Institute for the Neurosciences, Departments of Neurology & Psychiatry, Brigham & Women’s Hospital, Boston, MA; 2) Department of Neurology, Harvard Medical School, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA; 5) Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago, IL; 6) Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard University, Cambridge, MA; 7) Laboratory of Environmen-tation, Emory University School of Medicine, Atlanta, GA.

The DNA methylome captures the transcriptional potential of a cell or tissue. Differential methylation of validated Alzheimer’s disease (AD) loci could influence their effect; however where methylation falls along the disease pathway is unclear. We examine four causal models to assess the role of methylation on the pathology of AD. We utilized data from two longitudinal cohorts, the Religious Order Study and Rush Memory and Aging Project. DNA methylation profiles were generated in samples of dorsolateral prefrontal cortex using Illumina HumanMeth450K beadsets. We analyzed CpG sites within 25 kb of 11 validated AD susceptibility genes. The outcomes of interest were a measure of neuritic plaque (NP) accumulation and pathologic diagnosis of AD. First, independent associations between CpGs and outcomes were assessed using linear (NP) and logistic (AD) regression. The four hypothesized models are: (1) CpG mediated association, (2) reverse causality, (3) independent associations and (4) SNP by CpG interaction. Both (1) and (2) were assessed using mediation analyses and (4) was assessed for multiplicative interaction followed by stratified analyses. Correction for multiple testing was done using the Benjamini-Hochberg method. A total of 719 subjects were included in the analyses. Nine CpGs across 5 genes (BIN1,3 CLU,2,4 MS4A6A,2,4 ABCA7,2 APOE,1) and (APOE,1) were associated with the outcomes. Together they explain 13.1% of the variability in the combined outcomes. Most interestingly, a strong interaction effect was seen with the CR1 CpG at the 5’ end of the gene, where a significant inverse association was observed for the risk allele rs6656401, whereas those with the risk allele rs6656401TTA there is an inverse association between methylation and outcome, indicating more methylation is associated with less NP and decreased AD. These observations suggest that, within known AD susceptibility genes, methylation is related to pathologic processes associated with AD and may play a role in influencing gene expression from susceptibility loci.

429F Airborne Particulate Matter Exposure Modifies the Canonical MAP-Kinase Pathway: From Methylytic Analysis to Biological Implications. J.J. Carmona1,2,3, T. Sofer1,2,5, L. Cantone1, B. Reif1,2,4, A. Maity2, J. Schwartz1,2,3, X. Lin5,6, A. Baccarelli1,2,3,4,5,9) 1) Laboratory of Environmental Epigenetics, Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA; 2) Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; 3) Exposure, Epidemiology, & Risk Program, Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA; 4) Program in Quantitative Genomics, Harvard School of Public Health, Boston, MA, USA; 5) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 6) Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, IT; 7) Department of Statistics, North Carolina State University, Raleigh, NC, USA; 8) Dana-Farber/Harvard Cancer Center, Boston, MA, USA; 9) Harvard/Massachusetts General Hospital Center on Genomics, Vulnerable Populations, and Health Disparities, Boston, MA, USA.

Background: Exposure to air particulate matter with an aerodynamic diam-eter < 2.5 m (PM2.5) is known to elevate blood markers of inflammation and increase cardiopulmonary morbidity and mortality. Major components of PM2.5 are Black Carbon (BC) due to traffic and sulfate from coal-burning power plants. DNA methylation is known to be sensitive to environmental toxins and to mediate environmental effects on clinical outcomes via regulation of gene expression. We hypothesize that exposure to air pollution components affects DNA methylation in blood leukocytes, in genes from inflammatory pathways. Methods: 141 males from the Normative Aging Study (NAS), residing in the Boston area, were selected. Leukocyte DNA samples were hybridized to the RefSeq 38K Promoter tiling array (Roche Nimble Gen) and analyzed using the BIOCARTA website. Forward stepwise Canonical Correlation Analysis was applied to identify specific genes in the pathways associated with the exposure. Results: Methylation was associated with BC exposure, adjusted for sulfate and age, measured with a sensor located on the roof of the Countway Library at the Harvard Longwood Campus. 30-days moving average values of BC and sulfate were calculated for each participant at his clinical visit date. Sulfate measurements were applied to identify specific genes in the pathways associated with the MAP-kinase and NFkB signaling pathways were identified using the BIOCARTA website. Forward stepwise Canonical Correlation Analysis was applied to identify specific genes in the pathways associated with the expo-sure. The MAP-kinase pathway consists of 84 genes. Our analysis identified 10 genes whose methylation was associated with BC exposure, adjusted for sulfate and age (p-value 0.01). The association analysis between sulfate and methylation in this pathway suggested 9 genes, but was not statistically significant (p-value 0.086), which is possibly due to low power. There was no evidence of association between air pollution and methylation in the NFkB pathway. Conclusion: The effects of air pollution may influence inflammatory outcomes via MAPK gene methylation. These results will be validated in a larger subset of men from the NAS cohort.

428T Array-based assay detects genome-wide 5-methylcytosine and 5-hydroxymethylcytosine in non-human primates and mice. R.S. Alisch1,2, P. Chopra1, L.A. Papale3, A.T.J. White3, A. Hatch3, P.H. Roseboom1, M. Brown1, S. T. Warren1, 1) Psychiatry, Univ. of Wisconsin School of Medicine - Madison, Madison, WI; 2) Human Genetics, Biochemistry, and Pediatrics, Emory University School of Medicine, Atlanta, GA; 3) Exposure, Epidemiology, & Risk Program, Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA.

Murine and non-human primates (e.g. rhesus monkeys) represent excellent model systems to study human health and disease conditions, especially in the brain. However, use of these model systems for genomic profiling studies is limited because most array-based tools have only been developed to survey the human genome. Here we present the optimization of a widely used human DNA methylation array, designed to detect 5-methylcytosine (5-mC), and show that non-human data generated using the optimized array reproducibly distinguishes tissue types within and between chimpanzees, rhesus, and mouse, with correlations near the human DNA level (R2 > 0.99). While using this assay to conduct a genome-wide methylation analysis of rhesus placental and fetal tissues reveals 6,102 differentially methylated loci with a majority analysis significantly overrepresented for developmental processes, restricting the analysis to oncogenes and tumor suppressors genes finds 125 differentially methylated loci, suggesting that rhesus placental tissue carries a cancer epigenetic signature. Further optimization of the assay to detect 5-hydroxymethylcytosine (5-hmC) finds highly reproducible 5-hmC levels within human, rhesus, and mouse brain tissue that is species-specific with a hierarchical abundance among the three species (human > rhesus >> mouse). Together, these data show that this array-based methyla-tion assay is generalizable to all mammals for the detection of both 5-mC and 5-hmC, greatly improving the utility of mammalian model systems to study the role of epigenetics in human health, disease, and evolution.
Epigenomic fetal programming: identifying genomic sites differentially methylated after exposure to maternal gestational diabetes and responsive to its treatment. A.A. Houde1,2, S.M. Ruchat1,2, C. Allard3, M.R. Perron4−5, J.P. Baillargeon5, J. St-Pierre2,4, D. Gaudet3,5, D. Brisson6, M.F. Hivert3,6,7, L. Bouchard1,2, 1) Department of Biochemistry, Université de Sherbrooke, Sherbrooke, QC, Canada; 2) ECOGENE-21 and Lipid Clinic, Chicoutimi Hospital, Saguenay, QC, Canada; 3) Department of Medicine, Université de Sherbrooke, Sherbrooke, QC, Canada; 4) Department of Pediatricians, Chicoutimi Hospital, Saguenay, QC, Canada; 5) Department of Medicine, Université de Montréal, Montréal, QC, Canada; 6) Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA; 7) Massachusetts General Hospital, Boston, MA.

Background: In utero exposure to gestational diabetes mellitus (GDM) is associated with increased lifelong susceptibility to obesity and metabolic disorders for the offspring. Recent evidences showed that epigenetic modifications may be involved in the metabolic health programming of the newborn exposed to GDM. Nevertheless, whether the treatment of GDM women (diet or diet+insulin) has an impact on DNA methylation levels has not been established.

Hypothesis: DNA methylation at specific gene locus in placenta and cord blood is affected by exposure to GDM and its treatment. Methods: Placenta and cord blood samples were obtained from 43 women: 14 with normoglycemia (NGT) and 29 with GDM treated with diet (n=16, GDM-D) or diet+insulin (n=13, GDM-I). GDM was diagnosed between weeks 24-28 of pregnancy according to WHO criteria. DNA methylation was assessed at >485 000 CpG sites using the Illumina HumanMethylation450 BeadArray. DNA methylation differentiations between the 3 groups were determined using ANCOVAs (adjustment for infant sex, gestational age, maternal BMI at 1 trimester of pregnancy and history of GDM (P<0.05)) and significance of pairwise comparisons was verified with Tukey’s test. Results: Women were on average age 29 years old, GDM-I women were slightly overweight at 1 trimester (BMI=27.1 kg/m2) in comparison to GDM-D and NGT (23.9 and 24.6 kg/m2 (P=0.05)). In placenta exposed to GDM, lower levels of methylation were observed at TOX2 (P=6.8x10^-7; PAI9=1.2x10^-6) and DPP6 (P=8.8x10^-6; PAI9=1.9x10^-5) compared to NGT-exposed placenta. At PLB1 locus, lower methylation levels were observed in placenta from GDM-D women (77.5%) in comparison to NGT vs 73.6% in GDM-I (P=9.6x10^-8) whereas levels from GDM-I (78.2%) were similar to those from women with NGT. Similarly, in cord blood, we observed that CpG sites near GNASAS (PAI9=8.5x10^-5) and MYH7B (PAI9=4.5x10^-5) were differentially methylated in offspring from GDM-D mothers and were similar in NGT and GDM-I groups suggesting that insulin treatment offsets the impact of GDM exposure at these loci. In cord blood, methylation levels at STC2 were lower in offspring from mother with GDM, treated either with diet or insulin (PAI9=4.0x10^-7). Conclusion: Our results suggest that exposure to GDM or its treatment may influence DNA methylation at specific locus in the placenta and cord blood. The choice of clinical management of GDM may therefore have long lasting effect on the offspring’s epigenome and metabolic health.


Interaction between the genome, epigenome and environment is critical for the development of complex disorders. Expression of the genome may be influenced by the environment by shaping epigenetic mechanisms. At CpG-SNP sites, the genetic factors such as SNPs converge with environmentally-influenced epigenetic marks such as CpG dinucleotides, sites for cytosine methylation. The average occurrence rate of SNPs at a CpG site is 10-fold higher than the overall SNP occurrence rate, thus CpG-SNPs are overrepresented in the human genome (Xie et al., 2009). CpG dinucleotides are 7-fold more abundant at SNP sites than expected (Tosno and Bell, 2003). Correlation analysis identified allele specific methylation (ASM) on 30% heterozygous SNPs, and found that up to 88% of ASM regions are dependent on the presence of CpG-SNPs (Kerkel et al., 2008; Shoeemaker et al., 2010). Some CpG-SNPs may be associated with a disease, and alterations of their methylation under environmental influences may be a critical factor affecting gene expression and contributing to disease vulnerability. Such CpG-SNPs may be located within DNAse I hypersensitivity sites (DHSs) and targeted by generic and sequence-specific regulatory factors (RF). The CpG-SNP hypothesis received support in recent studies by us and others (John et al., 2011; Kaminsky et al., 2011; Martin-Trujillo et al., 2011; Reynard et al., 2011; Taqi et al., 2011; Urmi et al., 2011). We address this hypothesis by comparing abundance of CpG-SNPs and non-CpG SNPs at DHSs and binding places for 161 RFs in a panel of 91 cell lines (ENCE mode). We compared SNPs associated with alcohol dependence, behavioral disorders, neurodegenerative disorders, and the rest of SNPs implicated by NHGRI GWAS Catalog. As a model gene we analyzed methylation of CpG-SNPs in prodynorphin (PDYN), coding for opioid dynorphin peptides. Three PDYN-SNPs were significantly associated with alcohol dependence and differentially methylated in human brain. In the brain of alcoholics, methylation of the C allele of SNP rs2335749 (3'-UTR; C>T; C allele, risk variant) was increased (P<0.001) and positively correlated with alcohol intake. We also showed that methylation of these SNPs differently targets the T, risk allele, and the methylated and unmethylated C allele of this SNP, was identified. This findings support CpG-SNP hypothesis. Supported by Swedish FAS, VR and FORMAS. Tissues were received from the NSWTRC, University of Sydney.

Methylation analysis in tongue tissue of BWS patients identifies the (epi)genetic cause in 3 patients with normal methylation levels of H19 and KCNQ1OT1 in blood. M. Alders1, S.M. Maas1, D.M. Kadouch2, K. van der Lip1, H.J. Bleik1, C.M.A.M. van der Horst1, M.M.A.M. Mannens1. 1) Clinical Genetics, Academic Medical Center, Amsterdam, Netherlands; 2) Paediatrics, Academic Medical Center, Amsterdam, The Netherlands; 3) Plastic and Reconstructive Surgery, Academic Medical Center, Amsterdam.

Beckwith Wiedemann Syndrome (BWS) is caused by aberrant imprinting of genes in the chromosome 11p15 region. A molecular cause for BWS is found in approximately 80% of patients while 20% remains unexplained. We analyzed the imprinting status of H19 and KCNQ1OT1 in resected tongue tissue of 11 BWS patients, three of which had normal methylation levels in blood. In 8 patients with known methylation defects in blood (6 isolated KCNQ1OT1 hypomethylation, 2 UPD), the same defects were detected in tongue. The methylation levels of H19 in tongue were comparable to those found in blood. The methylation levels of KCNQ1OT1 were relatively higher in tongue than in blood in most patients. In all three patients with normal methylation levels in blood aberrant methylation patterns were found in tongue tissue. In two patients a UPD was detected and the third case had hypermethylation of H19. This result shows that tissue specific mosaic (epi)genetic changes, not present in blood, is the underlying defect in at least a subset of BWS patients without a molecular diagnosis after standard genetic testing.
433F
An Integrated epigenomic-transcriptomic Genetic analysis of schizophrenia brain identifies novel molecular pathways to disease. J. Mill², R. Pidgley², J. Viana¹, A. Jeffries³, C. Wong³, C. Troakes³, L. Schalkwyk¹, Exeter University, Exeter, Devon, United Kingdom; 2) Institute of Psychiatry, King’s College London, London, United Kingdom.

Schizophrenia (SZ) is a common psychiatric disorder characterized by the presence of psychotic symptoms and altered cognition. Although SZ is highly heritable, the molecular etiology of the disease is largely unknown. In addition to genetic and structural genomic variation, recent evidence supports a role for altered epigenomic and transcriptomic processes in disease pathogenesis. Frontal cortex and cerebellum tissue was obtained from 23 schizophrenia patients and 24 healthy controls. Genome-wide DNA methylation, expression and SNP profiling were performed using the Illumina Infinium Human Methylation450, HumanHT-12v4 Expression, HumanOmnimExpress BeadChips respectively. Integrated multi-level analyses provide evidence of SZ-associated DNA methylation and gene expression changes at biologically relevant loci, including GABBR1, RASA3, C8A, NRN1, BNP3, GAD1 and SERPINA3. Furthermore we identify cis-eQTLs and cis-mQTLs at SZ candidate gene names from published GWAS analyses, an increased burden of CNVs in patients with SZ, and a rare NRXN1 deletion in an SZ patient that is associated with altered DNA methylation. Together these results provide important insights into the biological mechanisms underlying SZ and highlight the value of taking an integrated ‘omics’ approach to complex disease.

434T
Whole Genome Bisulfite Sequencing of Cell Free DNA and its Cellular Contributors Links Placenta Hypomethylated Domains to Gene Deserts. T. Jensen¹, S. Kim¹, C. Chinn², Z. Zhu², T. Liu², C. Deciu², D. van den Boom², M. Ehrich², 1) Research and Development, Sequenom Center for Molecular Medicine, San Diego, CA; 2) Research and Development, Sequenom, San Diego, CA.

Circulating cell free (ccf) DNA is useful for non-invasive diagnostic testing in prenatal health and oncology. In both cases, the nucleic acid of interest is the minority species and thus needs to be differentiated from the highly abundant ccf DNA background. DNA methylation can serve as a method for distinguishing these; however, this depends on an in depth knowledge of the DNA composition. Whole genome bisulfite sequencing (WGBS) was performed on a set of unmatched samples including ccf DNA from 8 non-pregnant (NP) female donors, genomic DNA from 7 buffalo coat and 5 placenta samples, and ccf DNA from 7 pregnant females to gain a comprehensive understanding of the ccf DNA methylation in pregnant plasma. We first created a methylation map of ccf DNA from non-pregnant donors at single base resolution. Consistent with previous work in differentiated cell types, almost all cytosine methylation in NP ccf DNA samples occurred in the CpG context. We also found CpG cytosines within longer fragments were more likely to be methylated, linking DNA methylation and fragment size in ccf DNA. Next, we performed a series of pairwise comparative analyses to identify differentially methylated regions (DMRs). Comparison of the methylomes of placenta and NP ccf DNA enabled the detection of greater than 50,000 DMRs, the majority resulting from placenta hypomethylation. We found that >90% of these DMRs were located outside of CpG islands and were often associated with distinct histone tail modifications. Further investigation of the identified DMRs revealed the presence of large domains exhibiting consistent hypomethylation in placenta samples relative to NP ccf DNA across millions of consecutive bases. We found these domains to occur primarily in Cpg poor gene deserts. DMRs identified when comparing placenta to NP ccf DNA were recapitulated when comparing pregnant ccf DNA to NP ccf DNA, confirming the ability to detect differential methylation in ccf DNA mixtures. Overall, these data enabled the generation of a comprehensive map for each sample type at single base resolution, identified a link between local DNA methylation and ccf DNA fragment length, provided comprehensive lists of DMRs between sample groups, and uncovered the presence of megabase-size placenta hypomethylated domains. Furthermore, we anticipate these results to provide a foundation to which future studies to build non discriminatory DNA methylation for non-invasive testing can be compared.

435F
DNA Methylation at CPT1A is Associated with Triglyceride Levels, BMI and WHR. D.M. Absher¹, M.R. Irwin², S. Aslibekyan³, J. Sha³, L.L. Waite³, D. Zhu³, K. Stanton Thibeaul³, J. Ordovas³, D.K. Arnett³, 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) Department of Epidemiology, University of Alabama, Birmingham, Birmingham, AL; 3) Department of Biostatistics, University of Alabama, Birmingham, Birmingham, AL; 4) Department of Epimediology, University of Alabama, Birmingham, Birmingham, AL.

Epigenetic variation is thought to be a contributor to complex traits, and is likely to account for some of the missing heritability for phenotypes that have been incompletely explained by genetic variants. As the epigenome is dynamic and environmentally responsive, epigenetic modulators of dietary traits and responses are likely to be critically important to the development of cardiovascular disease and type-2 diabetes. We have undertaken an epigenetic analysis of GOLDN (Genetics of Lipid Lowering Drugs and Diet Network) study participants to identify DNA methylation patterns that contribute to lipid metabolic traits and related phenotypes. Using the Illumina Methylation450 array to measure DNA methylation at ~470,000 CpGs in CD4+ T-cells from 995 individuals in 183 families, we fit mixed effects regression models to identify CpGs where DNA methylation levels were strongly associated with baseline triglyceride levels (TG), body mass index (BMI), and waist-hip-ratio (WHR). We identified a CpG in the CPT1A (carnitine palmitoyltransferase 1A) gene that was significantly associated with all three traits (p=5.8e−28 for TG, p=2.8e−11 for BMI, and p=4.7e−11 for WHR), with ~2.4% of baseline TG variance explained by methylation at this locus. Given that CPT1A is an important regulator of fatty acid metabolism, we hypothesized that the impact of methylation on BMI and WHR would be mediated through effects on TG. Our findings support TG as a cpg-regression models using TG as a cpg-1 associated CpG near RPS6KA2 was significantly associated with BMI exclusively, including CpGs near the CD38, AHRG and PHDGH genes. Further, 1 additional CpG near RPS6KA2 was significantly associated with WHR.

436T
DNA Methylation Profiling is Robust in Different Tissue Types and Reveals Distinct Patterns Across Rheumatoid Arthritis Samples and Phenotypes. L.F. Barcellos¹, X. Shao¹, E. Elboudwarej¹, A. Baker¹, E. Sinclair¹, L.A. Criswell¹, 1) Div Epidemiology-SPH, Univ California, Berkeley, Berkeley, CA; 2) Div Rheumatology, Dept Medicine, Univ California, San Francisco, CA; 3) The UCSF-GIVI CFAR Immunology Core, Univ California, San Francisco, CA.

Rheumatoid arthritis (RA) is a chronic inflammatory disease with potential to cause substantial disability, primarily due to the erosive and deforming process in joints. RA etiology is complex with contributions from genetic and non-genetic factors. Epigenetic changes such as altered patterns of DNA methylation (DNAm), are also present in RA. Our goal was twofold: (1) to establish a protocol for performing DNAm profiling of specific immune cell populations isolated from large numbers of samples that is cost and labor efficient, and both accurate and reproducible; (2) to characterize DNAm in RA samples. We performed on a set of unmatched samples including ccf DNA from 8 non-pregnant (NP) female donors, genomic DNA from 7 buffalo coat and 5 placenta samples, and ccf DNA from 7 pregnant females to gain a comprehensive understanding of the ccf DNA methylation in pregnant plasma. We first created a methylation map of ccf DNA from non-pregnant donors at single base resolution. Consistent with previous work in differentiated cell types, almost all cytosine methylation in NP ccf DNA samples occurred in the CpG context. We also found CpG cytosines within longer fragments were more likely to be methylated, linking DNA methylation and fragment size in ccf DNA. Next, we performed a series of pairwise comparative analyses to identify differentially methylated regions (DMRs). Comparison of the methylomes of placenta and NP ccf DNA enabled the detection of greater than 50,000 DMRs, the majority resulting from placenta hypomethylation. We found that >90% of these DMRs were located outside of CpG islands and were often associated with distinct histone tail modifications. Further investigation of the identified DMRs revealed the presence of large domains exhibiting consistent hypomethylation in placenta samples relative to NP ccf DNA across millions of consecutive bases. We found these domains to occur primarily in Cpg poor gene deserts. DMRs identified when comparing placenta to NP ccf DNA were recapitulated when comparing pregnant ccf DNA to NP ccf DNA, confirming the ability to detect differential methylation in ccf DNA mixtures. Overall, these data enabled the generation of a comprehensive map for each sample type at single base resolution, identified a link between local DNA methylation and ccf DNA fragment length, provided comprehensive lists of DMRs between sample groups, and uncovered the presence of megabase-size placenta hypomethylated domains. Furthermore, we anticipate these results to provide a foundation to which future studies to build non discriminatory DNA methylation for non-invasive testing can be compared.

Mean DNAm levels were highly correlated between all cell types in both cases and controls; however, CD4+ memory T cells differed more from CD14+ monocytes (r=0.94) DNAm levels within putative RA genes identified through epigenetic analysis of GOLDN (43 genes; 883 CpG sites) were similar between cases and controls; however, some differences were noted. Hypermethylation in RA case B cells compared to controls for HLA-DRB1, CCR6, AFF3 and TAGAP CpG sites was observed, whereas, PTEN22 CpG sites were hypomethylated in case CD4+ naïve T cells compared to controls. Visualization of CpG DNAm profiles using multidimensional scaling revealed separation between cell types; results also suggest age and clinical phenotypic associations with DNAm. Our results emphasize the cell specific nature of DNAm, which has important implications for understanding human health and disease.
Gene networks for social cognition in Williams syndrome. L. Dai1, R. Weiss2, J.R. Korenberg1. 1) Center for Integrated Neuroscience and Human Behavior, Brain Institute, Department of Pediatrics, Univ. of Utah, Salt Lake City, UT; 2) Human Genetics, Univ. of Utah, Salt Lake City, UT.

Williams syndrome (WS), a neurodevelopmental disorder with hypersocial behavior, results from a deletion of ~28 genes on 7q11.23, that ultimately disturbs the neural circuitry involving oxytocin and vasopressin. Although cognitive deficits and social-emotional features are ultimately due to the deleted genes, the critical downstream pathways are unknown. Previous studies used rare WS genetic events, partial deletions seen in <2% of WS, to examine methylation differences. In this study, we took advantage of the social behavior, and showing that perturbations of transcriptional networks (34 WS patients and 18 typical controls), implicate disturbed function at synapse and dendritic spine. Here we address the possible role of epigenetic mechanisms in WS using genome-wide analyses of DNA methylation. We employed the Illumina Infinium HumanMethylation450 array to query the methylation of ~500K CpG/non-CpG sites in whole genome using whole blood DNA from 7 WS and 5 parents. The results revealed striking epigenetic changes throughout the genome, with differences from controls in 15,987 methylation sites (p<0.05), corresponding to 7478 genes. 72% of the significantly changed methylated sites are in CpG islands of which 67.1% are hyper-methylated in WS. Hyper and hypomethylated sites coexist within single genes. Further, the data indicates a possible correlation with parental origin and gender: hyper-methylation in maternal deletions (74.6%) vs 25.4% in paternals. Methylation also varies with gender, hypermethylation seen more in males (78.6%) than in females (21.4%), suggesting that the subtle differences in WS genes dramatically alter the epigenetic landscape of the genome, largely through imprinting mechanisms but also through high control.

Results Twenty CpG-SNPs in 6 genes, specifically, CCDC170, C7orf76, C9orf164, ESR1, SOX6 and WNT16, were differentially methylated in WS compared to controls. To identify novel pathways involved in the disturbed development and behavior seen in WS.

Identification of CpG-SNPs Associated with Osteoporosis. H. Shen1, C. Qiu1, L. Zhang1,2, C. Xu1, H.W. Deng1,2. 1) Center for Bioinformatics and Genomics, Department of Biostatistics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA; 2) Center of Systematic Biomedical Research, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

Background Osteoporosis is an increasingly serious public health problem, affecting over 100 million people worldwide. Osteoporosis is mainly characterized by low bone mineral density (BMD), which can lead to fractures. To date, over 50 BMD-associated loci have been identified, however, all these loci together account for no more than ~5% of BMD variation. In addition, despite the markers identified, specific functional variants are still unknown. In addition, the results of gene expression studies show that genome-wide DNA methylation profiles using Illumina HumanMethylation450 BeadChips in minor salivary gland biopsy tissue, fresh and banked peripheral blood mononuclear cells (PBMC), CD14+ monocytes, CD19+ B cells, and CD+ T cells in ten participants (five cases and five controls) from the Sjogren’s International Collaborative Clinical Alliance (SICCA; http://sicca.ucsf.edu/1) N01 DE023635 repository. Additionally, we characterized full genome SNP profiles. Sorting of freshly collected blood samples was performed using MACS®Technology. DNA yields for all cell and tissue types were high; all samples with foreground subtracted and normalized with beta-mixture quantile (BMQ). Methylation of individual measured CpG sites was stable across the six cell types for ~240,000 CpGs (variance< 0.001); however, variance > 0.10 for ~500 CpGs. Mean methylation was calculated within each cell type for the 473,929 CpGs detected by the BeadChip. Mean methylation levels varied between 0.1 to 0.99 for BMQ controls and 0.1 to 0.99 for all other cell types. We performed correlation analysis and in a combined data set (n= 12: 0.51-0.99) with the lowest correlation between salivary gland tissue and CD14+ monocytes and the highest correlation between fresh and banked PBMCs. Methylation levels within 17 putative BS genes identified with GWAS were similar between cases and controls; however, median methylation levels for BLK and TNFSF4 CpGs were 10%–18% less in cases compared to controls within salivary gland tissue. Data visualization of methylation profiles using Multi-Dimensional Scaling revealed distinct separation between cell types in addition to separation based on case status with the greatest separation based on case status visible in CD19+ B cells and salivary gland tissue samples. Our results emphasize the cell and tissue specificity of DNA methylation. Additional research, including studies of tissue-specific expression, and loss of function or gain of function will be required to fully define the role of DNA methylation in SS.
The results indicate the presence of a mosaic subpopulation of epigenetically-dysregulated, ectodermally-derived cells in subjects with ASD. The epigenetic dysregulation observed in these ASD subjects born to older mothers may be associated with aging parental gametes, environmental influences during embryogenesis or could reflect mutations of the chromatin regulatory genes increasingly implicated in ASD. The results indicate that epigenetic dysregulatory mechanisms may complement and interact with DNA mutations in the pathogenesis of the disorder.

1) Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany; 2) Division of Molecular Medicine, Institute Ruer Boskovic, Zagreb, Croatia; 3) Institute of Human Genetics, University of Würzburg, Germany; 4) Laboratory of Molecular Biology, National Institute of Mental Health, Bethesda, MD.

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

**444T**

DNA differential methylation is observed at BRCAl promoter but not in 8q24.21 in cleft lip and palate. L. Alvizi1, G.S. Kobayashi, C.B.F. Silva, D.Y. Sun1,2, R.S. Passos-Bueno. Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Sao Paulo, Brazil. Purpose: DNA methylation is known to be a heritable regulatory mechanism in gene expression and influenced by both genetic and environmental factors. It is also known that impairment in gene methylation status may lead to gene expression dysregulation and thus disease. In this context, cleft lip and palate (CL/P) is a congenital craniofacial malformation with high incidence (1,700 live births) strongly determined by the genetic and environmental interplay in which epigenetic factors such as DNA methylation are very plausible factors in the malformation etiology. Aiming to investigate DNA methylation at specific sites to CL/P, we investigated whether BRCAl, previously associated to CL/P (Kobayashi and Alvizi et al, 2013 PloS ONE), and 8q24.21 CL/P risk region were differentially methylated in CL/P patients in comparison to control samples. Methods: Bisulfite sequencing analysis for BRCAl promoter was performed in a DNA sample set obtained from dental pulp stem cells (DPSC) of 18 CL/P and 12 controls and for 8q24 region from white blood cells DNA of 34 CL/P and 44 controls. A total of 300 clones for BRCAl promoter and 780 clones for 8q24.21 were sequenced and analysis was performed using BISMA (Bisul- fite Methylation Analysis - BPCD online tool). BRCAl expression was also assessed by qRT-PCR in the DPSC sample. Results/Conclusions: Total BRCAl promoter methylation was significantly higher (+1.4%) in the DPSC CL/P sample. Besides, BRCAl promoter CpGs 1, 2 and 11 were the most hypermethylated in the CL/P sample (17.8%, 30.2% and 23.1%, respectively). As expected, BRCAl expression was significantly reduced in comparison to controls (p=0.001). No evidence of differential methylation at the 8q24.21 cleft lip risk locus was found in the white blood cells DNA sample of CL/P patients as compared to controls. Our results suggest that downregulation of BRCAl in CL/P samples may be driven by increased BRCAl promoter methylation and the causative factors in this hypermethylation should be next investigated. BRCAl expression rescue by promoter demethylation is being conducted in DPSC CL/P samples. FAPESP/CNPq-MCT.
447F
Methylation QTLs often show opposite allelic directions when comparing different tissues. M.J. Bonder¹, S. Kasela¹,², K. Kirotar³, M. Kals³, M. Ivanov⁴, A. Metspalu¹,², M. Ingelman-Sundberg⁴, C. Wijmenga¹, A. Zhernakova¹, L. Milani², L. Franke³. ¹) Genetics, University Medical Center Groningen, Groningen, Groningen, Netherlands; ²) Estonian Genome Center, University of Tartu, Tartu, Estonia; ³) Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; ⁴) Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden.

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

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

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

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

448T
Predicting Prostate Cancer Progression through Gene Network Analysis of Methylation Data. L. Briollais, K. Kron, B. Bapat, H. Oezcelik. Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, M5T 1L9, Canada.

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

449F
A pilot study testing DNA methylation profiles in Samoan obese and lean young adult males. O.D. Buhule¹, N.L. Hawley², M. Medvedovic², R.L. Minster², G. Sun², H. Cheng², S. Vissio², R. Deka², D.E. Weeks², S.T. McGarvey³. ¹) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ²) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ³) Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio 45267, USA; 4) Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH, USA; 5) Weight Control and Diabetes Research Center, The Miriam Hospital, Providence, RI, USA & The Alpert Medical School, Brown University, Providence, RI, USA; 6) Medical Specialist Clinic and National Health Services, Government of Samoa, Apia, Samoa; 7) International Health Institute and Department of Epidemiology, Brown University School of Public Health, Providence, RI 02912, USA.

Background and Objective: Methylation levels, which influence gene expression, can be influenced by environment and life style, Obesity, as a product of both nutritional environment and life style, could be related to methylation levels. Here we present preliminary findings from a pilot study examining DNA methylation patterns across the genome in young obese and lean male Samoans to identify epigenetic loci associated with obesity.

Methods: DNA was extracted from whole peripheral blood from 46 obese (BMI >=32 & Abdominal Circumference >=92.5cm) and 46 lean (BMI <26 & Abdominal Circumference <92.5cm) men. The DNA was bisulphite converted and then amplified using the primer pairs designed in our previous study (Ivanov et al, PLoS Genetics 2012, Fairfax et al, Nature Genetics 2012). Here we present preliminary findings from a pilot study examining DNA methylation patterns across the genome in young obese and lean male Samoans to identify epigenetic loci associated with obesity.

Results: DNA was extracted from whole peripheral blood from 46 obese (BMI >=32 & Abdominal Circumference >=92.5cm) and 46 lean (BMI <26 & Abdominal Circumference <92.5cm) men. The DNA was bisulphite converted and then amplified using the primer pairs designed in our previous study (Ivanov et al, PLoS Genetics 2012, Fairfax et al, Nature Genetics 2012). Here we present preliminary findings from a pilot study examining DNA methylation patterns across the genome in young obese and lean male Samoans to identify epigenetic loci associated with obesity.

Support: This work was supported by NIH R01 HL093093 (PI: Stephen McGarvey).
Contribution of DNA methylation to gene expression varies by tissue and age. C. Chen1,2, C. Zhang1,2, L. Cheng1,2, J. Badner1, E. Gershon1, J. Sweeney1, R. Badilla-Don1, B. Don1, D. T. Butcher1, D. Grafodatskaya1, D. W. X. Wei1, W. Reardon1, B. B. McCusker1, A. Verloes1, F. J. Flam1, R. Popescu2, R. Bastille-Porras3, R. Mendoza-Londono4, R. Weksberg1,6,8,1, Genetics & Genome Biology, SickKids Research Institute, Toronto, Ontario, Canada; 2) National Centre for Medical Genetics, Our Lady’s Children’s Hospital, Dublin, Ireland; 3) 6th Department of Otolaryngology, The Hospital for Sick Children, Toronto, Canada; 4) 6th Department of Otolaryngology, University of Texas Southwestern, Medical Center, Dallas, TX; 5) Department of Psychiatry, Northwestern University, Chicago, IL; 6) 6th Department of Otolaryngology, The Hospital for Sick Children, Toronto, Canada; 7) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada; 8) Department of Paediatrics, The University of Toronto, Toronto, Canada.

CHARGE syndrome (CHARGE) is a rare autosomal dominant genetic disorder, with an incidence of 1 in 8500–10000 births. Clinical diagnosis for CHARGE is based on non-random associations of the following congenital abnormalities: Coloboma of the eye, Heart defects, Atresia of the choanae, Retarded growth and development, Genital abnormalities, Ear abnormalities/deafness/vestibular disorder. In the majority of cases, CHARGE is the result of haplosufficiency due to loss of function, missense, or deletion in the gene encoding Chromodomain Helicase DNA-binding protein (CHD7). Targeted mutational studies in Drosophila (kismet) and mouse (Chd7) have found phenotypes similar to those found in human. In Drosophila reduced expression of kismet/CHD7 results in deficits in axonal pruning, guidance and extension, as well as defects in memory and learning function. Normal mammalian growth and development depend on the correct epigenetic programming of the genome. Epigenetic patterns evolve across development utilizing mechanisms such as DNA methylation and covalent modifications of histones and non-coding RNA. In Drosophila, kismet ortholog of human CHD7 has been demonstrated to regulate the repressive histone H3 methylation mark of lysine 27. In human cell lines, CHD7 has been shown to bind to chromatin regions that are active as demonstrated by histone H3 lysine 4 methylation and DNA hypermethylation of these binding sites. CHD7 also interacts with RNA polymerase II, forming complexes that alter chromatin structure to facilitate access for transcriptional machinery. These epigenetic modifications of histone H3 are tightly linked to DNA methylation patterns. We hypothesized that specific DNA methylation alterations occur as a result of the heterozygous CHD7 mutations and could reveal critical downstream targets associated with CHARGE clinical features. We have analyzed cases with CHD7 mutations comparing their methylation alterations to age and sex-matched controls using the Illumina Infinium Methylation450 BeadChip array. Data were analyzed using the iMA package in R and Genome Studio software from Illumina. We identified both gain and loss of methylation in genes that play a role in growth and neurodevelopment. The identification of these epigenetic modifications could lead to an improved understanding of the pathophysiology of CHARGE and the type of chromatin regions to which CHD proteins are recruited.
A specific DNA methylation signature associated with NSD1+/− mutations in Sotos syndrome reveals a significant genome-wide loss of DNA methylation (DNAm) targeting CGs in regulatory regions of key developmental genes. S. Choufani1, C. Cytrynbaum2, A.L. Turinsky2, Y.A. Chen3, D. Grafovatskaya4, J. Xiang5, M. Feigenberg5, B.Y.H. Chung6, D.J. Stavropoulos7, R. Mendoza-Londono8, D. Chitayat9, W.T. Gibson9, M. Reardon9, M. Bruder10,11, R. Weksberg10,11,12 1 Program in Genetics and Genome Biology, Hosp Sick Children, Toronto ON, Canada; 2) Div Clin & Metabolic Gen, Hosp Sick Children, Toronto, ON, Canada; 3) Molecular Structure & Function, Hosp Sick Children, Toronto, ON, Canada; 4) Centre for Computational Medicine, Hosp Sick Children, Toronto, ON, Canada; 5) Dept of Paediatrics & Adolescent Med, Li Ka Shing Faculty of Medicine, Hong Kong; 6) Paediatric Laboratory Medicine, Hosp Sick Children, Toronto, ON, Canada; 7) Dept of Medical Genetics, UBC, Child and Family Research Institute, Vancouver, BC, Canada; 8) Our Lady's Hospital for Sick Children, Crumlin, Dublin 12, Ireland; 9) Department of Computer Science and Donnelly Centre, University of Toronto, Toronto, ON, Canada.

Sotos syndrome (SS) is characterized by somatic overgrowth and intellectual disability. Most SS cases (>75%) have mutations in NSD1 (nuclear receptor-binding SET domain protein 1). NSD1 binds near promoter elements and regulates transcription initiation and elongation via interactions with H3-K36me and RNA polymerase II. To determine if NSD1 mutations impact stable epigenetic marks such as DNA methylation (DNAm), we compared DNAm in peripheral blood DNA from SS cases with NSD1 mutations (NSD1+/−; n=20) to controls (n=30) using the Illumina Infinium450methylation BeadChip. Differential DNAm analysis using non-parametric statistics (with correction for multiple testing) coupled with permutation analyses identified a surprising number of targets (n=2157) with differences in DNAm (DM CG sites with >20% difference in DNAm) between SS and controls. These sites were distributed across the genome; 95% demonstrated loss of DNAm. Using unsupervised hierarchical clustering of the 2157 DM CG sites, all SS cases with NSD1+/− clustered as a distinct group separate from controls. Moreover, DNAm at these sites clearly distinguished SS (NSD1+/−) from Weaver syndrome (EZH2+/-, n=5), another overgrowth syndrome which has been noted to have methylation changes. We further identified 4,541 CG sites that are unmethylated in SS and demonstrate intermediate levels of methylation in lung tissue and finally 9,830 sites where there is complete methylation in lung tissue but no methylation in blood tissue. Next we evaluated specific genes which have been associated in genomewide association studies with lung related diseases. We noted good correlation between blood and lung methylation values at ORMDL3 asthma locus but poor concordance at the Thymic Stromal Lymphopoietin (TSLP) locus. Conclusion: We have identified 49,376 CpG sites with tissue specific methylation, including CpG sites in the asthma gene ORMDL3.

Comparison of methylation profiles in human blood and lung tissue identifies tissue specific CpG methylation sites. D. Daley1, K. Ushey2, L. Akhabir3, A. Saferai3, S.M. Mah4, A. Sandford5, M.S. Kodor6, R. Paré7 1) James Hogg Research Center Department of Medicine University of British Columbia Vancouver, BC, Canada V6G Y6; 2) Department of Medical Genetics University of British Columbia Vancouver, BC, Canada V5Z 1H4.

We hypothesize that differences in methylation patterns between individuals may contribute to the etiology of asthma, COPD and other lung related traits. As the majority of genetic studies collect and maintain blood samples from large participant cohorts, participants to be included in the study must be available. Our approach is to use tissue specific methylation (TSM) patterns between blood and lung tissues. Methods: To evaluate TSM we used 36 paired samples (blood and lung tissue from the same individual) and 22 lung only samples. Methylation of over 450 thousand CpG sites throughout the genome was evaluated using the Illumina Infinium Human Methylation450K bead chip array. Methylation levels were assessed using beta values which range from 0 (no methylation) to 1 (complete methylation). Principal components were used to identify 3 samples which clustered differently, all 3 samples were noted to have low and/or poor quality DNA and were removed from subsequent analyses. Correlations between beta values were examined using principal components, heatmaps, and a mixture model of three distributions (unmethylated, variable methylation, complete methylation). Results: We identified 49,376 CG sites with apparent tissue specific methylation patterns. These sites are 31% enriched with binding sites in 41% of the targets (NCOR=H3K36me3 marks in both normal blood and embryonic stem cells. Also, mapped to enhancers and CpG island shores. Analysis of ChIP-seq data (n=1) and non-pathogenic (n=5) variants. The majority of these DM CG sites constitute a DNAm signature that is specific for NSD1+/−. Moreover, DNAm at these sites clearly distinguished SS (NSD1+/−) from Weaver syndrome (EZH2+/-, n=5), another overgrowth syndrome which has been noted to have methylation changes. We further identified 4,541 CG sites that are unmethylated in lung tissue, and demonstrate intermediate (variable) methylation patterns in SS tissue. Comparison of methylation between blood and lung tissue sites with complete methylation in blood and intermediate levels of methylation in lung tissue and finally 9,830 sites where there is complete methylation in lung tissue and intermediate levels in blood tissue. Next we evaluated specific genes which have been associated in genomewide association studies with lung related diseases. We noted good correlation between blood and lung methylation values at ORMDL3 asthma locus but poor concordance at the Thymic Stromal Lymphopoietin (TSLP) locus. Conclusion: We have identified 49,376 CG sites with tissue specific methylation, including CpG sites in the asthma gene TSLP.


Alcohol abuse during pregnancy can lead to a range of neurological abnormalities termed Fetal Alcohol Spectrum Disorder (FASD). The mechanisms by which alcohol (ethanol) induces FASD are poorly understood; however, disruption of the regulation of genetic programs in the brain may be involved. We have previously developed a model to generate FASD-like mice showing characteristic learning and memory impairment. Mouse pups were injected with saline or ethanol on postnatal days 4 and 7, the period equivalent to human trimester three. At 70 days of age, hippocampi were isolated and used for gene and miRNA expression microarray, and methylated DNA immunoprecipitation & microarray (MeDIP-chip). MeDIP-chip was used to assess changes in histone H3 lysine 4 methylation (H3K4me3) and histone H3 lysine 27 methylation (H3K27me3). Fifty nine genes were differentially expressed at p<0.05 and fold cut off 1.2. Ingenuity Pathway Analysis (IPA) identified the top affected biological pathway as ‘Free Radical Scavenging, Gene Expression, Dermatological Diseases and Conditions’ with 13,400 genes affected. Five genes had gene expression and DNA methylation changes; one such gene was Mafg, downregulated by 1.21 fold. MAFG is a transcription factor involved in oxidative stress response. Two miRNAs which target Mafg, miR-134a and miR-200b were upregulated. Mafg methylation as a result of miRNA and DNA methylation changes induced by ethanol. Ethanol is known to induce oxidative stress in the developing brain through a variety a mechanisms including reduction of antioxidant levels and increasing reactive oxygen species (ROS) production. Alternatively, the epigenetic regulation of Mafg and other oxidative stress response genes may represent a novel point of interface between the epigenetic and oxidative stress mechanisms of FASD generation.
456T

Genome-wide DNA methylation profiles in fruit flies and the effect of huntingtin knockout. S. Erdin1, K. Dietz1,2, A. Ragavendran1, M.E. Talkowski1,2, J.A. Walker3,4, J.F. Gusella4,5 1 Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Louisiana State University Health Sciences Center, Shreveport, LA; 3) Department of Neurology, Harvard Medical School, Boston, MA; 4) Department of Genetics, Harvard Medical School, Boston, MA; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA.

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

458T

Epigenetic changes in relation to asbestos exposure in malignant pleural mesothelioma. G. Fiont1,2, S. Garrera1, E. Casalone1,2, M. Bett3, E. Aldieri4, D. Ferrante5, C. Di Gaetano5,2, F. Rosai6, A. Russo6,7,8, S. Tunets9, M. Padano9, A. Asperti9, C. Casadio9, A. Ardissone9, E. Ruffini9, P.G. Bett9, R. Libener9, R. Guaschin9, E. Piccilli9,10, D. Mirabelli11,12, C. Magnani11,12, I. Dianzani13, G. Matullo1,2. 1) Human Genetics Foundation, HùDeP, I-10126 Turin, Italy; 2) Department of Medical Sciences, University of Turin, I-10100, Turin, Italy; 3) Laboratory of Genetic Pathology, Department Health Sciences, University of Piemonte Orientale, I-28100, Novara, Italy; 4) Department of Oncology, University of Turin, I-10126 Turin, Italy; 5) CPO Piemonte and University of Medical Statistics and Epidemiology, Department Translational Medicine, University of Piemonte Orientale, I-28100, Novara, Italy; 6) Thoracic Surgery Unit, University of Piemonte Orientale, I-28100, Novara, Italy; 7) Chest Surgery, Department of Clinical and Biological Sciences, University of Turin, I-10043, Orbassano, Italy; 8) Thoracic Surgery Unit, University of Turin, I-10126, Turin, Italy; 9) Pathology Unit, Azienda Ospedaliera Nazionale SS, Antonio e Biagio e Cesare Arrigo, I-15121, Alessandria, Italy; 10) Transfusion Centre, Azienda Ospedaliera Nazionale SS, Antonio e Biagio e Cesare Arrigo, I-15121, Alessandria, Italy; 11) Pneumology Unit, Santo Spirito Hospital, I-15033, Casale Monferrato, Italy; 12) Unit of Cancer Epidemiology, CPO-Piemonte and University of Turin, I-10126, Turin, Italy; 13) Interdepartmental Center for Studies on Asbestos and other Toxic Particulates ‘G. Scansetti’, University of Turin, I-10125, Turin, Italy.

Malignant pleural mesothelioma (MPM) is a rare and aggressive tumor strongly associated with asbestos exposure. Only 5–17% of individuals exposed to asbestos develop MPM, suggesting the involvement of other environmental, genetic and epigenetic risk factors. DNA methylation is a critical mechanism of gene silencing in human malignancies. The relationship between aberrant DNA methylation and inflammation has been documented in many types of cancers, including MPM. Asbestos exposure may contribute to MPM onset through this relationship. We conducted an epigenome-wide scan to identify differentially methylated regions (DMR) in 40 MPM cases versus 40 controls, and in asbestos high-exposed versus low-exposed subjects. Methylation status was measured for about 470k CpG sites in DNA from blood and MPM cell lines of MPM patients. Illumina data confirm the presence of methylated DNA sites in Drosophila melanogaster and suggest a significant role of huntingtin in epigenetic modification.

457F

A novel method for identification and quantification of consistently differentially methylated genomic regions. C. Fann1,2, C.L. Hsiao1, C.J. Chang2. 1) Epidemiology & Genetics, Inst Biomed Sci, Acad Sinica, Taipei, Taiwan; 2) Graduate Institute of Clinical Medical Science, Chang Gung University, Taoyuan, Taiwan.

Advances in biotechnology have resulted in large-scale studies of DNA methylation. A differentially methylated region (DMR) is a genomic region with multiple adjacent CpG sites that exhibit different methylation statuses among multiple samples. Many so-called ‘supervised’ methods have been established to identify DMRs between two or more comparison groups. Methods for the identification of DMRs without reference to phenotypic information are, however, less well studied. An alternative ‘unsupervised’ approach was proposed, in which DMRs in studied samples were identified with consideration of nature dependence structure of methylation measurements between neighboring probes from tiling arrays. Through simulation and further downstream analyses are ongoing, but these data confirm the presence of genome-wide gender-specific differences, and suggest differential methylation associated with loss of huntingtin. Replication of these findings and further development of these methods will be possible.
Genetic Ancestry Explains Differences in Local and Methylation Patterns in the GALA II Study. J.M. Galanter1, C.R. Gignoux2, S.S. Oh1, D.G. Torgerson1, C. Ying1, S. Huntsworth1, R.Merchant3, D. Hu1, S. Shen1, M. Pino-Yanes1, E. Nguyen4, P. Avila1, H.J. Farber1, A. Davis5, E. Birgino-Buenaventura6, M.A. Lenoir6, K. Meade7, D. Serebrisky8, T. Rossi9, W. Rodriguez-Chintron10, R. Kumar11, J.R. Rodriguez-Santana11, E.G. Burchard1. 1) University of California, San Francisco San Francisco, CA; 2) Feinberg School of Medicine, Northwestern University, Chicago, IL; 3) Baylor College of Medicine and Texas Children’s Hospital, Houston, TX; 4) Children's Hospital and Research Center Oakland, Oakland, CA; 5) Kaiser Permanente-Vallejo Medical Center, Vallejo, CA; 6) Bay Area Pediatrics, Oakland, CA; 7) Jacobi Medical Center, Bronx, NY; 8) San Francisco General Hospital, San Francisco, CA; 9) Veterans Carribean Health System, San Juan, PR; 10) The Ann and Robert H. Lurie Children’s Hospital of Chicago, Chicago, IL; 11) Centro de Neumologia Pediatrica, San Juan, PR.

Epigenetic modification of the genome through methylation plays a key role in the regulation of diverse cellular processes. Changes in DNA methylation patterns have been associated with many complex diseases. Recent studies have found significant differences in the methylation patterns of peripheral blood between African Americans and non-Hispanic Whites. In this study, we leveraged estimates of genomic ancestry in 575 Latino children of multiple Latino ethnicities (Puerto Rican, Mexican, and other) enrolled in the GALA II study of childhood asthma to determine whether differences in global and local methylation patterns between ethnic groups could be explained by ancestry. We measured DNA methylation at ~450,000 markers using the Illumina Infinium HumanMethylation450 BeadChip. We used multidimensional scaling to determine global methylation patterns. Ethnicity was significantly associated with at least six principal components of ancestry (P < 0.001). We performed a mediation analysis to determine the extent to which genomic ancestry mediated the effect of ethnicity on local methylation. Of the 316 methylation sites with a p-value less than 1 × 10−10, 40 (13%) were significantly mediated by Native American ancestry. The median proportion of the effect of ethnicity on PC6 and for the comparison of Puerto Ricans to Mexicans). Native American ancestry, when added to the model, was also highly associated with the principal coordinate, and its inclusion in the model explained the significance of the association between ancestry and the methylation pattern measured by PC6. We then performed an epigenome-wide association study between ethnicity and methylation at each site. There was a significant association between ethnicity and local methylation patterns at 176 sites (at a Bonferroni corrected significance level of 1 × 10−7). We performed a mediation analysis to determine the extent to which genomic ancestry mediated the effect of ethnicity on local methylation. Of the 316 methylation sites with a p-value less than 1 × 10−10, 40 (13%) were significantly mediated by Native American ancestry. The median proportion of the effect of ethnicity on methylation mediated by Native American ancestry was 66% across all sites. An epigenome-wide association study between Native American Ancestry and methylation found 309 sites with a Bonferroni corrected significance level of 1 × 10−7 or below. These findings have broad implications for the study of methylation patterns across populations and for disease association studies. There is significant differences in methylation patterns between ethnic groups that are due to ethnicity differences in those groups. These differences should be accounted for in performing epigenome wide disease association studies.

Edematous severe childhood malnutrition is associated with widespread DNA hypomethylation. N.A. Hanchard1, K. Marshall2, X. Wang3, T.E. Forrestere, M.E. Reide1, J.W. Belmon4, C.A. McKenzie5, 1) USDA/ARS/Children's Nutrition Research Center, Baylor College of Medicine, 1100 Bates Ave, Houston, TX, 77030, USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX, 77030, USA; 3) Tropical Metabolism Research Unit, Tropical Medicine Research Institute, University of the West Indies, Mona, Kingston 5, Jamaica.

Severe childhood malnutrition (SCM) is a major global health problem that contributes to more than two million childhood deaths worldwide each year. SCM occurs in two clinically distinct forms -- the more severe edematous SCM (ESCM), and the milder non-edematous SCM (NESC). The etiology of this clinical dichotomy is unclear; however, previous studies have observed differences between acutely ill SCM and NESC patients in the flux of methyl-groups through the 1-carbon conversion of methionine to homocysteine. Because this reaction contributes significantly to the maintenance of CpG dinucleotide methylation during mitosis, we hypothesized that in high-turnover tissues, DNA methylation in ESCM patients might differ from that in NESC patients. To test this hypothesis, we used the Illumina Infinium HumanMethylation450 BeadChip array to evaluate methylation at ~485,000 genome-wide CpG sites in buccal DNA samples collected from Jamaican children recruited shortly after recovery from acute SCM, and Jamaican adults recruited 18 years after having SCM. After data normalization and quality control, linear regression models were used to determine CpG sites with differential methylation between ESCM and NESC patients. Immediately post-SCM recovery, we found that samples from ESCM children (N=9) showed significant global hypomethylation relative to NESC (N=8) at 21,000 CpG sites genome-wide (false discovery rate, FDR, p<0.01), involving 1,461 genes with more than two significant probes. In pathway analyses, the top 1,000 differentially hypomethylated probes were 4-fold enriched for genes involved in GTPase regulator activity (FDR, p<1 × 10−3). Conversely, there was no statistically significant differences in methylation between ESCM (N=10) and NESC (N=14) individuals sampled as adults. Our results provide evidence for genome-wide differential DNA methylation in ESCM that is directly related to the nutritional insult, and provide a basis for larger scale studies that will illustrate our observations. Further, these results suggest that methylation may be a useful intermediate phenotype with which to probe possible genetic trait loci that could provide insight into why some children develop ESCM while others develop NESC, despite similar environmental and dietary exposures.
463F

Fibromyalgia (FM) is a debilitating condition associated with multiple symptoms, including chronic, widespread pain; dyscognition; fatigue; sleep disturbances; and depression. These symptoms often stem from a decreased workplace productivity, increased health/disability costs, and compromises in the overall quality of life for people with FM. While the adverse health and economic impacts of FM have been well recognized, its etiology remains enigmatic. We hypothesized that FM is associated with the development of acquired somatic epigenetic and/or chromosomal alterations. To test this hypothesis we compared the frequency of spontaneously occurring micronuclei (MN) and genome-wide methylation patterns (illumina 450K HumanMethylation Chip) in leukocytes of women with FM (N=10) to those seen in comparably aged healthy controls (N=42 [MN]; N=8 [methylation]). A 3.26-fold significantly increased mean frequency of MN (t2 =45.552; df = 1; p=1.49x10^-11) was observed in the patients with FM [mean=51.4, sd= 21.9] when compared to the controls [mean=15.8, sd=8.5]. Significant differences (N=69 sites) in methylation patterns were also observed between cases and controls (5% false discovery rate), with 91% of these differentially methylated regions (DMRs) being attributable to increased methylation values in the women with FM. The DMRs involved 15 significant biological functions (including but not limited to) genes involved in neuron differentiation, nervous system development, skeletal/organ system development, and chromatin compaction. Specific genes with a DMR(s) included the brain-derived neurotrophic factor (BDNF) gene (noted to play an important neurotrophic regulatory role in pain transduction, learning and memory), as well as genes involved in muscle contraction (NR4A3; HDAC4; FEZ1; PKG1), chromatin compaction (NAT15; HDAC4; UHRF1), and DNA damage or chromosomal segregation (SOD3; UHRF1; NAT15). In summary, these results support the role of epigenetic and acquired chromosomal alterations as a possible biological mechanism leading to or resulting from symptoms associated with FM. Ultimately, since epigenetic changes demonstrate plasticity, recognition of consistent epigenetic alterations associated with FM could provide a means for development of therapeutic or intervention strategies to treat this debilitating condition.

464T
Epigenome-wide methylation patterns across multiple fetal tissues. A.C. Just1, A.A. Baccarelli2, R.J. Wright3, H.H. Burns1, G. Estrada1, L. Schnaas4, R.O. Wright2. 1) Harvard School of Public Health, Boston, MA, USA; 2) Icahn School of Medicine at Mount Sinai, NY, USA; 3) Department of Neonatology, Beth Israel Deaconess Medical Center, Division of Newborn Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA; 4) National Institute of Perinatology, Mexico City, Mexico.

Rationale: Methylation is a common epigenetic mark with regulatory functions important for cell differentiation, normal development, and disease. Methylation is also responsive to environmental exposures. New tools enable measureable changes in methylation with single-base resolution across the genome which can be utilized in large epidemiologic studies. A limitation however is that typically only surrogate tissues such as white blood cells have been used. Human genomic methylation patterns in many potential target tissues have not yet been characterized. Objectives: To compare genomic methylation across multiple fetal tissues (cord blood, placenta, umbilical vein, umbilical artery) collected from the same individuals and identify similar and distinct patterns across samples. Methods: All four fetal tissues were collected for DNA extraction from participants (n=9, total tissue samples=36) in Mexico City. Samples were run on the Illumina Human Methylation 450K beadchip and reporting % methylation values (betas) were adjusted with subset quantile normalization (Touleimat and Tost 2012). Methylation profiles were compared using multi-dimensional scaling and Pearson’s correlation analysis using all probes passing QC as well as subsets defined based on genomic context. Measurements and Main Results: After initial quality control 428,056 CpG sites passed filtering for all 36 samples. Sub-setting to the normal probes, the largest Pearson’s correlations between tissues, calculated for each participant, were for umbilical vein and umbilical artery (range 0.975 to 0.995) and the lowest for cord blood and placenta (range 0.796 to 0.834) with similar values when restricted to sites located in CpG islands. Conclusions: Correlation analysis demonstrates that DNA from umbilical vein and artery have more similar patterns of methylation than cord blood versus placental DNA. In a larger cohort analysis currently underway (n = 130 participants), methylation measures in these fetal tissues will be related back to prenatal exposure to inhaled environmental stressors (e.g., lead, air pollution, psychosocial stress) and as mediators of associations between exposures and birth outcomes.

465F
Mapping the DNA unmethylose with mTAG, a novel method to investigate the epigenome. V. Labriola1, E. Knukienė2, T. Khare1, G. Urbanavicius3, A. Lapinaite2, K. Končevičius2, D. Lītā1, T. Wang4, S. Pai1, C. Ptak5, J. Gordoževičius3, S.C. Wang6, A. Petronis1, S. Klimašauskas1. 1) The Krembil Epigenetics Laboratory, Centre for Addiction and Mental Health, Toronto, Canada; 2) Department of Biological DNA Modification, Institute of Biotechnology, Vilnius University, Vilnius, Lithuania; 3) Faculty of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania; 4) Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, U.S.A; 5) Institute of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania; 6) Institute of Systems Biology and Bioinformatics, National Central University, Chungli, Taiwan.

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

466T
Data analysis for identifying differentially methylated regions. M. LeBlanc1,2, C. Page1,2, A. Frigessi1,2, B. Kulle Andreassen1,2. 1) Epi-Gen, Institute of Clinical Medicine, Akershus University Hospital, University of Oslo and Akershus University Hospital, Oslo, Norway; 2) Department of Biostatistics, University of Oslo, Oslo, Norway; 3) Department of Neurology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

There is general agreement that methylation studies need to report results for differentially methylated regions (DMR). As such, methylation studies are well suited for the application of statistical methods that simultaneously analyze all available data for a given region of interest. One approach for this is functional data analysis (FDA), where statistics are calculated using functions describing the pattern of methylation over a defined region. In contrast, most methylation studies to date either use statistics based on single methylation probes or summarize single probe values within a region of interest with one value (e.g. maximum or mean). There are also methods using various smoothing methods, where the statistical approach operates with a summary measure based on the smoothed values (e.g. Jaffee 2012). We apply statistical methods based on FDA to methylation data. Using simulated data, we explore the utilities of FDA in the context of methylation and compare it to other methods used in the literature. In addition, we apply the proposed method to a recent dataset. The results encourage the use of FDA for methylation data in predefined regions of interest.
467F
DNA methylation analysis of iPS cells using whole-genome bisulphite sequencing. D. Lee1,2, J. Shin1, P. Tonge4, M. Purf1, A. Nagy4,6, J. Seo1,2,3,1
1) GMI, Seoul national university, Seoul, South Korea; 2) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea; 3) Macrogen Inc., Seoul 153-023, Korea; 4) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto M5S 1AB, Canada.

Somatic cell reprogramming involves epigenetic remodeling of chromatin architecture including DNA methylation, conferring induced pluripotent stem cells (iPSCs) with characteristics similar to embryonic stem (ES) cells. However, it remains unclear how the epigenetic pattern changes during the reprogramming process and how it controls the messenger RNA expression. To address these questions, we utilized a secondary inducible reprogramming system developed in mouse using piggyBac-mediated transposition of the four reprogramming factors. Here we report whole-genome DNA methylation profiles at single-base resolution for mouse ES cells, secondary mouse embryonic fibroblast (MEF), secondary factor independent iPSCs induced from these MEFs, and cells undergoing reprogramming from Day 2 to 18, using whole genome bisulftite sequencing, along with comparative analysis of RNA expression and Histone modification (H3K4me3, H3K27me3, H3K36me3) using NGS. We clarified how the epigenetic change controls RNA expression and showed CpG methylation blocks the activation of core pluripotency genes. We defined 7990 differentially methylated regions (DMRs) in CpG context. Ectopic expression of reprogramming factors leads to a dynamic CpG methylation change from start of the process. Gain of CpG methylation occurs rapid and massive around PRCs (RING1, Suz12, EZH2) binding sites both in ESC like and aberrant direction. CpG methylation of transcription factor binding sites of 4 factors (Oct4, Sox2, Klf4, cMyc) or the factors activated early in reprogramming (Nanog) showed demethylation during high dox treatment, but only at the exact binding sites. However the samples became pluripotent showed demethylation in wide region around binding sites of all the core pluripotency factors, including Esrrb, Tcflp21, and Zfx. We also observed that H3K4me3 and H3K27me3 are biased to regions showing low CpG methylation. These regions show dynamic change of both histone modifications and RNA expression. Thus, the regions with high CpG methylation are restricted to non-Histone modification and the genes having high CpG methylation promoters showed low expression. Our data provide evidence for how ectopic reprogramming factors affect CpG methylation in early stage of reprogramming and what is essential to gain pluripotency.

468T
Correlation and Null Hypothesis in Epigenome-wide Association Studies (EWAS). B. Lehne1, A. Drong2, M. Loh1, W. Zhang1, W. Scott1, M.R. Jarvelin1,4,6, P. Elliott1,2, M.I. McCarthy1,7, J.S. Koonen7, J.C. Chambers7,1
1) Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 3) MRC-HPA Centre for Environment and Health, Imperial College London, London, United Kingdom; 4) Institute of Health Sciences, University of Oulu, Oulu, Finland; 5) National Institute of Health and Welfare, Oulu, Finland; 6) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 7) National Heart and Lung Institute, Imperial College London, London, United Kingdom.

DNA methylation plays an important role in the regulation of gene expression and may be involved in the molecular mechanisms that lead to many human diseases. Currently multiple large-scale Epigenome-wide Association Studies (EWAS) are in progress to investigate the relationships between DNA methylation and phenotypic variation. In this work we analyse the p-value distribution that arises in an EWAS under no association, using permutation testing of Infinium 450K data for 2,660 individuals. Under no association we observe a substantial deflation of test statistics (median genomic inflation factor \(\lambda_{\text{median}}=0.92\)). We demonstrate that this apparent deflation is caused by close correlation between methylation markers, even if these markers have entirely different genomic loci (mean pairwise correlation \(\bar{r}=0.23\) for 1,000 randomly selected markers). This correlation structure substantially reduces the number of independent tests and affects the distribution of test statistics under the Null Hypothesis.

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

We conclude that correlation between markers is the consequence of biological and technical confounders, each of which affect the methylation status of multiple markers simultaneously. Correlation between markers affects the Null Hypothesis underlying an EWAS. Analysis of EWAS data is enhanced by careful adjustment for these confounding factors.
469F  Importance of Batch and White Blood Cell Subtypes Correction in Analysis of Illumina Infinium 450K Methylation Arrays. M. Loh1,2, B. Lehnert1, A. Bronn1, W. Scott1, M. Zhang1, M.-R. Jarvelin3,4, P. Elliott1,4, M.I. McCarthy1,2, J.S. Kooper1, J.C. Chambers1,1. 1) Epidemiology and Bio-statistics, Imperial College London, London, United Kingdom; 2) Institute of Health Sciences, University of Oulu, Oulu, Finland; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 4) MRC-HPA Centre for Environment and Health, Imperial College London, London, United Kingdom; 5) National Institute of Health and Welfare, University of Oulu, Oulu, Finland; 6) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom.

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

471F  Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. C. Nagy1,2, M. Suderman3,4, M. Sztyf5, N. Mecha-1,2, C. Enroth6,7, G. Verhaeghe8, J. Szyf1,2. 1) Department of Pharmacology and Therapeutics, McGill University; 2) McGill Centre for Bioinformatics; 3) McGill Centre for Suicide Studies, Douglas Mental Health University Institute; 4) McGill Centre for Bioinformatics; 5) Department of Pharmacology and Therapeutics, McGill University, Montreal, Quebec, Canada; 2) McGill Group for Suicide Studies, Douglas Mental Health University Institute; 3) McGill Centre for Bioinformatics; 4) Department of Pharmacology and Therapeutics, McGill University, Montreal, Quebec, Canada; 2) McGill Group for Suicide Studies, Douglas Mental Health University Institute; 3) McGill Centre for Bioinformatics; 4) Department of Pharmacology and Therapeutics, McGill University, Montreal, Quebec, Canada.

Background: Methylome abnormalities in the brain are known to be involved in depression and suicide. Open field tests in mice revealed that social withdrawal and anxiety-like phenotypes are associated with astrocytic changes in the brain. To date, the epigenetic consequences of astrocytic alterations in the brain have not been described. Astrocytes play a central role in nervous system activity and control synaptic function by releasing and reabsorbing the neurotransmitters glutamate and glycine. The expression of astrocyte markers such as GFAP, ALDH1L1, SOX9, GLUL, GLYT2, and SLC1A3 is associated with depression and suicide, providing a potential framework for better understanding of these behaviors.

Conclusions: The known functions of affected genes suggest that the affected genes included those previously implicated in psychosis. Also, the affected genes included those previously implicated in psychosis.

472T  The effects of perinatal testosterone exposure on DNA methylation in the brain are late-emerging and dynamic. T.C. Ngun1,2, N.M. Ghahramani1,2, P.Y. Chen1,2, S. Krishnan1,2, S. Muir1,2, T. TeSlaa1, L. Rubbi2, A.P. Arnold3,4, G. de Vries5, N. Forger6, M. Pelligrini7, E. Villain1,2, 1) Department of Human Genetics, David Geffen School of Medicine at University of California Los Angeles (UCLA), Los Angeles, CA, USA; 2) Laboratory of Neuroendocrinology of the Brain Research Institute, UCLA, Los Angeles, CA, USA; 3) Department of Molecular, Cellular, and Developmental Biology, UCLA, Los Angeles, CA, USA; 4) Department of Integrative Biology and Physiology, UCLA, Los Angeles, CA, USA; 5) Neuroscience Institute, Georgia State University, Atlanta, GA, USA.

We hypothesized that perinatal exposure to T will affect DNA methylation patterns in two sexually dimorphic brain regions - the bed nucleus of the stria terminalis/preoptic area (BNST/POA) and the striatum - and that this exposure contributes to the establishment of sex differences in the methylene. We established the first genome-wide methylation profiles for males (XY), females (XX), and females treated with T at birth (XX+T) at two time points. The first was postnatal day 4 (PN4) (within the critical window for T’s organizational effects) and the second was PN60 (adulthood). Our approach was validated by the finding that the majority of X chromosome CpG sites differing between XX and XY mice at PN60 were hypermethylated in XX animals in both regions of the brain, consistent with X chromosome inactivation. The short-term methylome exposure was relatively modest: 45 genes were differentially methylated between XX and XX+T in the BNST/POA, while 68 genes were influenced by T in the striatum. However, by PN60, this number had grown dramatically. The number of T-influenced genes was 760 in the BNST/POA and 1377 in the striatum. The traditional view of brain organization that T plays a critical role in brain development is, however, not supported by our findings. Our approach was validated by the finding that only a few of the genes affected by T at PN4 were the same as those affected at PN60 in either region was unexpected. Methylom at sexually dimorphic CpG sites in XX+T was more masculinized at PN60 than PN4 in both regions. However, the masculinizing effect of T on the methylene was stronger in the striatum than in the BNST/POA. Taken together, these data suggest that the organizing effect of T on the methylene is late-emerging and dynamic. Our data also suggest that organization by T may occur via early programming that is critical to the development of the brain to respond in a particular fashion to later developmental events. Drs. Ngun and Ghahramani contributed equally to this work.
**473F**

**Genome-wide DNA Methylation Network Analysis for Osteoporosis**

**Risk.** c. qiu1, H. Shen1; J. Li1; H.W. Deng1.2. 1) Center for Bioinformatics and Genomics, Department of Biostatistics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA; 2) Center of Systematic Biomedical Research, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

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

---

**474T**

**Genome-wide placental DNA methylation analysis of severely growth-discordant monozygotic monochorionic twins reveals a novel epigenetic signature for IUGR.** M. Rolfman1, S. Choufani2, S. Drewlo3, S. Keating4, J. Kingdom5, R. Weksberg1,2. 1) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 2) Genetics and Genome Biology Program, Hospital for Sick Children Research Institute, University of Toronto, Toronto, Canada; 3) Research Centre for Women’s and Infants’ Health, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Canada; 4) Department of Laboratory Medicine and Pathology, Mount Sinai Hospital, University of Toronto, Toronto, Canada; 5) Maternal-Fetal Medicine Division, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada.

**Background:** Intrauterine growth restriction (IUGR) refers to reduced fetal growth in the context of placental insufficiency. IUGR is not only associated with perinatal morbidity and mortality, but has also been linked to adult-onset diseases, such as obesity, hypertension and diabetes, and thus poses a major health burden. Placental epigenetic dysregulation has been implicated in IUGR; however, clear pathophysiological mechanisms remain to be elucidated. Monozygotic monochorionic (MZ-MC) twins are particularly affected by IUGR in the setting of severe discordant growth with unequal placental share. Because MZ twins have the same genotype at conception, and a shared maternal environment, they provide a unique model for studying epigenetic regulation of the placenta (i.e. the non-shared environment). **Objective:** We compared genome-wide placental DNA methylation patterns of severely growth-discordant MZ-MC twins to identify novel candidate genes for IUGR in the context of unequal placental share. **Methods:** We studied 8 severely growth-discordant MZ-MC twin pairs, each with unequal placental share. Snap frozen placental samples were obtained at delivery from each twin. Placental pathology was histologically determined. We used the Infinium HumanMethylation450 BeadChip array platform to identify genomic regions exhibiting differential methylation in IUGR versus normal twins. **Results:** Our analysis of larger twins (n=8) versus smaller twins (n=8) identified 138 candidate genes with statistically significant (p<0.05) differential methylation (using a cutoff of 10% difference in DNA methylation). Across these 138 genes, the lower weight twins exhibit a distinctly different methylation signature compared to their higher weight counterparts. These genes, many linked to IUGR for the first time in this study, confer a variety of functions, including cellular development, signaling, angiogenesis, and lipid metabolism. Largest methylation differences between the two groups were found in the DECR1, LEPR and SPG7 genes, implicating lipid metabolism and mitochondrial dysfunction as prominent pathophysiological mechanisms in this form of IUGR. **Conclusion:** We propose a novel epigenetic signature for IUGR in the context of unequal placental share in MZ-MC growth-discordant twins, highlighting lipid metabolism and mitochondrial dysfunction as major contributors, and suggesting an underlying mechanism for the fetal reprogramming associated with adult-onset diseases.
475F Genetics of global DNA methylation patterns in multiple tissues from twins. J.K. Sandling, E. Grundberg, E. Meduri, A.K. Hedman, S. Keilsson, N. Nielsen, A. Barrett, K.S. Small, B. Ge, E.T. Dermitzakis, M.I. McCarthy, T.D. Spector, J.T. Bell, P. Deloukas, the MuTHER consortium. 1) Wellcome Trust Sanger Institute, Hinxton, UK; 2) Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Churchill Hospital, Oxford, UK; 5) McGill University and Genome Quebec Innovation Centre, Montreal, Canada; 6) Department of Genetic Medicine and Development and Institute for Genetics and Genomics in Geneva, University of Geneva Medical School, Geneva, Switzerland; 7) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK; 8) Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 9) Present address: Department of Human Genetics, McGill University, McGill University and Genome Quebec Innovation Centre, Montreal, Canada; 10) Present address: Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK.

Epigenetic mechanisms, such as DNA methylation, are attractive candidate processes for explaining the interplay of genetic and environmental factors in complex traits. We explored the associations between genetics and epigenetics utilizing genome-wide DNA methylation profiles (Illumina HumanMethylation450) from skin (N=469) and adipose tissue (N=648) from female twins in the MuTHER study. We found that methylation levels were generally highly correlated between skin and adipose tissue from the same individual ($r_{median}=0.96$); this dropped significantly if only the (10%) most strongly tissue-dependent sites were considered ($r_{median}=0.71$). For around 5% of CpG sites we observed marked differences in average methylation levels between the tissues ($\Delta beta=0.3$). We found methylation variability to be suppressed in regions known to impact gene regulation such as promoters (adipose: $P<1E-5$, skin: $P<1E-4$). We then studied the degree of association between adipose DNA methylation and expression of nearby genes in this tissue, and of 13,532 genes we found 17% to have at least one significant association with a slight enrichment of negative correlations. Concordance in methylation for the same CpG site in skin tissue was generally lower (r~0.4) than in adipose tissue (~0.6) and this dropped significantly if only the (10%) most strongly tissue-dependent CpG sites were considered ($r_{median}=0.3$). By age 35% of methylation levels were predicted to vary significantly between skin and adipose tissue.


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

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

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

477F Whole-genome bisulfite sequencing (WGBS): A novel ‘post-bisulfite conversion’ library construction method from low qDNA inputs. R. Sooknanan, J. Hitchen, D. Gabel, V. Ruotti. Epicentre Biotechnologies, LLC (an Illumina® company), Madison, WI, USA mina® company, Madison, WI.

Genome-wide analysis of 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) are possible with whole-genome bisulfite sequencing (WGBS), where unmethylated cytosine residues are converted to uracil. Further, with the use of T4 β-glucosyltransferase (T4-BGT) and Ten-eleven Translocation Gene Protein 1 (Tet1) enzymes, 5-hmC can be discriminated from 5-mC on a genome-wide basis. However, a major challenge in WGBS is the degradation of DNA that occurs during bisulfite conversion under conditions required for complete conversion. Typically, ~90% of input DNA is degraded and thus, is especially problematic with limited starting amounts of DNA. Additionally, regions that are rich in unmethylated cytosines are more sensitive to strand breaks. As a consequence, a majority of DNA fragments contained in di-tagged NGS DNA libraries treated with bisulfite ‘post-library construction’ can be rendered inactive due to strand breaks in the DNA sequence flankned by the adapter sequences. These mono-tagged templates are then excluded during library enrichment resulting in incomplete coverage and bias when performing whole genome bisulfite sequencing. Here, we describe a novel ‘post-bisulfite conversion’ library construction method for preparing NGS libraries from genomic DNA prior to the addition of the adapters. This ‘post-bisulfite conversion’ library construction method uses the untagged single-stranded DNA as template for the subsequent addition of adapter sequences required for NGS. Thus, single-stranded DNA fragments independent of size and position of strand breaks remain as viable templates for library construction, eliminating the loss of adapter fragments and the selection bias associated with a ‘post-library construction’ bisulfite conversion strategy. This novel ‘post-bisulfite conversion’ library construction method exhibits increased sensitivity and efficiency and improved coverage required for WGBS for detecting 5-mC and 5-hmC marks.
478T


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

479F

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

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

480T

Epigenome-wide differences in DNA methylation of autosomes in Klinefelter’s Syndrome. E.S. Wan1, W. Qiu1, J. Morrow1, T.H. Beatty2, J. Hetmanski2, E.K. Silverman3, D.L. DeMeo1 on behalf of the COPIDGene Investigators. 1) Channing Division of Network Medicine, Brigham and Women’s Hospital, Boston, MA; 2) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

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

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

Results: There were no significant differences by race or current smoking status between KF subjects and controls; KF subjects were significantly younger and had a higher mean BMI than controls. 204 Infinium I sites and 16 Infinium II sites, including sites annotated to the G-protein coupled receptor 27 (GPR27) and Fas associated factor 1 (FAF1), met a priori threshold for significance.

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

481F

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

Ascorbate (vitamin C) is best known for its role in scurvy, in which the hydroxylation of collagen catalyzed by collagen hydroxylases is incomplete due to ascorbate deficiency. Recently, Tet (ten-eleven translocation) family dioxygenases (also known as Tet proteins) have been identified to hydroxylate 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in DNA. Both Tet proteins and collagen hydroxylases belong to the Iron and 2-oxoglutarate-dependent dioxygenase superfamily. These dioxygenases utilize Fe2+ as a cofactor, 2-oxoglutarate as a co-substrate, and some of them including collagen hydroxylases require ascorbate as another cofactor for full catalytic activity. We found that the content of 5hmC was extremely low in mouse embryonic fibroblasts (MEFs) cultured in ascorbate-free medium. Additions of ascorbate dose- and time-dependently enhanced the generation of ShmC, without any effects on the expression of Tet genes. Treatment with another reducer glutathione (GSH) did not change the level of ShmC. Further, blocking ascorbate entry into cells by phloretin and knocking down TET1-3 expression by short interference RNAs (siRNA) significantly inhibited the effect of ascorbate on ShmC. The effect of ascorbate on ShmC generation appears independent on iron uptake by cells or the production of 2-oxoglutarate. These results suggest that ascorbate enhances ShmC generation, most likely by acting as a cofactor for Tet proteins to hydroxylate 5mC to 5hmC. Thus, we have uncovered a novel role for ascorbate in modulating the epigenetic control of genome activity.
A new target enrichment system generating ultra-high complexity probe pools for targeted bisulfite sequencing, J. Wendt1, M. Suzuki2, T. Richmond1, N. Patterson1, T. Millard1, D. Groen1, R. Bannen1, R. Selzer1, T. Albert1, J. Jeddeloh1, J. Greally1, D. Burgess1. 1) Roche NimbleGen Inc., Development and Technology Innovation Groups, Madison, WI; 2) Albert Einstein College of Medicine, Center for Epigenomics, Bronx, NY.

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

Posters: Epigenetics

482T

Racial and Ethnic Variation in DNA Methylation of Human Peripheral Blood. Y. Yang1, R. Siebra1, Y. Li1, J.F. DeCoteau1, I. Peter1, S.A. Scott1. 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Anesthesiology, Xuzhou Medical College, Xuzhou, Jiangsu, China; 3) Cancer Stem Cell Research Group, University of Saskatchewan, Saskatoon, SK, Canada.

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

483F

Epigenome-wide association study of Autism Spectrum Disorder using Post-Mortem Brain Tissue. C.C.Y. Wong1, D.H. Geschwind1,2,3, N. Parkash4, C. Troakes1, J. Viana1, D. Condiffe1, T.G. Beltard1, S. Prabaharan1, J. Mills1,2,1) King’s College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, De Crespigny Park, London, UK; 2) Center for Autism Research and Treatment, and Program in Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA; 3) Department of Human Genetics, University of California, Los Angeles, CA, USA; 4) Program in Neurogenetics, Department of Neurology, University of California, Los Angeles, CA, USA; 5) King’s College London, MRC Centre for Neurodegeneration Research, Institute of Psychiatry, De Crespigny Park, London, UK; 6) University of Exeter Medical School, Exeter University, St Luke’s Campus, Exeter, UK; 7) Computational and Systems Biology, Genome Institute of Singapore, 60 Biopolis St, Singapore 138672, Singapore.

Autism Spectrum disorders (ASD) is a group of complex neurodevelopmental disorders characterized by considerable etiological heterogeneity. Although genetic variation in several dozen ASD genes has been implicated in the development of ASD, collectively accounting for 10-20% of ASD cases, there is emerging evidence to suggest a role for epigenetic variation in the disorder. In this study, we performed an epigenome-wide association study (EWAS) of ASD using a large collection of post-mortem brain tissue (n=138) obtained from four brain regions. DNA methylation was quantified using the Illumina Infinium 450K Human Methylation array in conjunction with an analysis pipeline developed by our group. In this study, we identified a number of significant disease-relevant differentially methylated regions (DMRs), including several located in the vicinity of genes previously implicated in psychiatric disorder. Pathway analysis of our top loci highlighted a significant enrichment of epigenetic disruption in biological networks and pathways relevant to disease and neurodevelopment. Overall, our data provide further evidence to support a role of differential DNA methylation in the etiology of ASD.
Gene based association analysis of brain DNA methylation with Alzheimer’s disease pathology using random permutation. L. Yu1, G. Sinivas- tava2,3,4, B. Chitnik2,3,4, M. Eaton3,4, S. Leurgans1, A. Meissner2,4, P.L. De Jager1,2,3,4, D.A. Bennett1. 1) Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago, IL; 2) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology & Psychiatry, Brigham and Women’s Hospital, Boston, MA; 3) Harvard Medical School, Boston MA; 4) Broad Institute of MIT and Harvard, Cambridge, MA; 5) Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA; 6) Department of Stem Cell and Regenerative Biology, Harvard University, Boston, MA.

Background: Gene based analysis of DNA methylation serves as a useful tool to test the global hypothesis for regional association of CpGs with traits of interest. In order to minimize the inflation of Type I error, the correlations between CpGs must be considered. However, the correlation structure of CpGs has not been fully understood. We propose a permutation based method to evaluate methylation associations by gene regions controlling for correlations among CpGs.

Methods: We analyzed gene based associations of brain DNA methylation at 11 validated AD loci, APOE, CR1, BIN1, CD33, CLU, ABCA7, CD2AP, PICALM, EPHA1, MS4A6A and MS4A4A, with the pathologic diagnosis of Alzheimer’s Disease (AD). We utilized data from two ongoing longitudinal cohort studies of aging and dementia, the Religious Orders Study (ROS) and the Memory and Aging Project (MAP). Genomic wide methylation profiles were generated from dorsolateral prefrontal cortex tissue samples using Illumina HumanMethylation450 beadset. We first applied logistic regression models with AD diagnosis according to NIA Reagan criteria as the binary outcome, and examined the associations of methylation with AD at the gene level after adjusting for covariates such as age, sex, race, batch effects and cell heterogeneity. Similar methods were used to test whether DNA methylation levels in saliva predicted those of blood. For all analyses, the FDR was controlled at 5 percent. Hierarchical clustering segregated blood samples from saliva samples. Overall, DNA methylation of only ~10 percent of CpGs from saliva predicted that of blood. While no individual CpG site remained associated with AD after adjustment for multiple testing, the test statistics for analyses of saliva and blood were moderately correlated (r=0.21). We evaluated CpG sites in 11 genes that were reported as differentially methylated in the blood of those with a history of child abuse. The majority of those genes contained CpG sites whose methylation levels were highly predictive of those in blood (p<10^-7) though certain areas of the gene (i.e. promoters) were more likely to contain such sites. These results have applications for longitudinal and biomarker studies as well as large-scale DNA methylation studies of childhood psychiatric disorders. DNA methylation derived from saliva may be informative for research questions that can be assessed in blood, though only a small fraction of CpG sites can be considered correlated. Tissue-specific differences seem more prominent than those related to shared genetic or environmental factors.

487F
The utility of DNA extracted from saliva for methylation studies of psychiatric traits. A.K. Smith1,2, V. Kilaru1, T. Klengel3, K.B. Mercer4, K. N. Connelly1,2, K.J. Ressler1, E.B. Binder1. 1) Psychiatry & Behavioral Sciences, Emory University, Atlanta, GA; 2) Max Planck Institute of Psychiatry, Munich, Germany; 3) Genetics and Molecular Biology Program, Emory University, Atlanta, GA; 4) Human Genetics, Emory University, Atlanta, GA. DNA methylation has become increasingly recognized as a fundamental factor in the etiology of psychiatric disorders. Because brain tissue is not accessible in living humans, epigenetic studies are often conducted in blood, but few are conducted in children because even a blood draw may be too invasive. Saliva DNA extraction is readily available in pediatric and adult settings and is associated with a high level of inter-individual reliability. In saliva DNA, DNA methylation is measured as a percent of methylated CpGs at a specific site.

Aim: To evaluate the utility of DNA extracted from saliva for methylation studies of childhood psychiatric disorders.

Methods: DNA methylation was interrogated with the HumanMethylation450 BeadChip. The proportion of epithelial cells in saliva and specific leukocyte types in blood were estimated. We examined the association of each CpG site with child abuse using linear models that adjusted confounding factors such as age, sex, race, batch effects and cell heterogeneity. Similar methods were used to test whether DNA methylation levels in saliva predicted those of blood. For all analyses, the FDR was controlled at 5 percent. Hierarchical clustering segregated blood samples from saliva samples. Overall, DNA methylation of only ~10 percent of CpGs from saliva predicted that of blood. While no individual CpG site remained associated with AD after adjustment for multiple testing, the test statistics for analyses of saliva and blood were moderately correlated (r=0.21). We evaluated CpG sites in 11 genes that were reported as differentially methylated in the blood of those with a history of child abuse. The majority of those genes contained CpG sites whose methylation levels were highly predictive of those in blood (p<10^-7) though certain areas of the gene (i.e. promoters) were more likely to contain such sites. These results have applications for longitudinal and biomarker studies as well as large-scale DNA methylation studies of childhood psychiatric disorders. DNA methylation derived from saliva may be informative for research questions that can be assessed in blood, though only a small fraction of CpG sites can be considered correlated. Tissue-specific differences seem more prominent than those related to shared genetic or environmental factors.

488T

Friedreich ataxia (FRDA) is caused by a homoyzgous GAA repeat expansion mutation within intron 1 of the FNX gene, leading to reduced expression of frataxin protein. Evidence suggests that the mutation induces epigenetic changes and heterochromatin formation, thereby impeding gene transcription. Thus, studies using FRDA lymphocytes and lymphoblastoid cell lines have detected increased DNA methylation of specific CpG sites upstream of the GAA repeat, together with histone acetylation and methylation changes in the region.

In this study, we performed an unbiased DNA methylation changes in FRDA brain, cerebellum and heart tissues, which are the primary affected systems of the disorder. Bisulfite sequence analysis of the FNX flanking GAA regions revealed a shift in the FRDA DNA methylation profile, with upstream CpG sites becoming consistently hypermethylated and downstream CpG sites becoming consistently hypomethylated. However, bisulfite sequencing is a time-consuming procedure. Therefore, we have now developed specific ‘methyl’chron sequencing and qPCR-based protocols to more rapidly quantify DNA methylation at four of the upstream CpG sites. Analysis of human cerebellum and heart tissue samples was used to validate the technique by comparison with previous bisulfite sequencing results. Increased DNA methylation in both FRDA cerebellar and heart tissue samples compared with unaffected samples was confirmed at all 4 CpG sites. Increased DNA methylation at all four CpG sites in the upstream GAA repeat region was also confirmed in human FRDA fibroblast cells, which have not previously been characterized. The levels of DNA methylation at two CpG sites in the upstream GAA repeat region correlated with transcripts that confer the FRDA phenotype in fibroblasts. DNA methylation and frataxin expression levels were determined in Y47R control and YG8R FRDA mouse fibroblasts, neural stem cells (NSCs) and neuronal/glial differentiated NSCs. In all three cell types, there is increased DNA methylation and decreased frataxin expression in Y47R FRDA cells compared with YG8R control cells. These changes correlate with the expression of a methyl ‘methyl’ cron in FRDA cells and tissues and may provide a potential target for FRDA treatment.

DNA methylation has become increasingly recognized as a fundamental factor in the etiology of psychiatric disorders. Because brain tissue is not accessible in living humans, epigenetic studies are often conducted in blood, but few are conducted in children because even a blood draw may be too invasive. Saliva DNA extraction is readily available in pediatric and adult settings and is associated with a high level of inter-individual reliability. In saliva DNA, DNA methylation is measured as a percent of methylated CpGs at a specific site.
489F Systematic classification of common disease-associated regulatory DNA variations by their epigenomic relationship. M. Duzmovic, C. Giles, J. Wren. Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK.

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

490T Integrated epigenetic analysis of APOBEC family gene in breast cancer. Y. Zhang, R. Delahanty, W. Zheng, J. Long. Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee.

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

491F Role of CTCF in epigenetic regulation of 4qD4Z4 macrosatellite repeat. J. Lim1, J.M. Moore2, N.A. Rababah2, Y.D. Krom2, S.J. Tapscott1,2,3. 1) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 3) Department of Neurology, University of Washington, Seattle, WA.

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

492T Subtelomeric deletions exert telomere position effects on nearby genes. J. Gerfen, H.M. Mason-Suarez, M.K. Rudd. Department of Human Genetics, Emory University. Atlanta, GA.

Copy number variations (CNVs) are a major cause of autism, intellectual disability, and birth defects. In some cases, loss or gain of genes within the CNV can explain clinical features in affected individuals. However, many pathogenic CNVs are associated with patient phenotypes without pinpointing the gene(s) involved. Position effect, where genes nearby but outside the CNV have altered expression, is another process that could be involved in disease etiology. In particular, telomere position effect (TPE) is a well-described epigenetic process in model systems where genes are silenced by proximity to a heterochromosomal telomere. We measured TPE in lymphoblastoid cell lines derived from patients with pathogenic subtelomeric deletions using ChIP-chip and RNA-seq followed by targeted analysis with qPCR and next-generation sequencing of PCR amplicons. Analysis of four different subtelomeric deletions demonstrated a variety of chromatin and expression changes in genes adjacent to deletions. In one 5.5-Mb deletion of the short arm of chromosome 4 we detected an ~400-kb region enriched in histone three lysine nine trimethylation, a marker of heterochromatin, directly adjacent to the breakpoint. RNA-seq analysis of EVC, a gene ~150-kb from the breakpoint and within the region of heterochromatin, revealed allele-specific skewing of expression. We found that one allele accounts for ~80% of the transcripts, suggesting that the EVC locus on the other chromosome is at least partially silenced. Other genes outside the region of heterochromatin were expressed at similar levels from both alleles. Silencing of EVC may be responsible for the atrial septal defect in this patient. Heterozygous mutations of EVC have been reported in a family with congenital heart defects, and recessive mutations lead to Ellis-van Creveld syndrome, which includes structural defects of the atria. These data support our hypothesis that heterochromatin originating from a new telomere can silence genes adjacent to a subtelomeric deletion. Other deletions in our study showed weaker TPE. We are currently characterizing additional chromatin modifications associated with human TPE and measuring the distance of TPE spreading in other CNVs. Our experiments are the first to demonstrate TPE from a native human telomere. These data will elucidate the role of CNV position effects in disease and may reveal new genes involved in patient phenotypes.
493F Chromatin state characterization of GWAS results of different neuropsychiatric traits is suggestive of brain-specific as well as non-neuronal origins of disease. R.A. Ophoff1, J. Ernst1. 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, California; 2) Deps of Biological Chemistry and Computer Science, UCLA, Los Angeles, California.

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

494T Histone H2A variants: one family, different roles in the human genome. M.Y. Tolstorukov1,2, J.A. Goldman1, C. Gilbert1, V. Ogryzko3, N. Volfovsky4, R.M. Stephens5, R.E. Kingston1, P.J. Park6. 1) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA; 2) Division of Genetics, Brigham and Women’s Hospital, Boston, Massachusetts, USA; 3) Oncogene, Differentiation and Transduction du Signal, Institut Andre Lwoff, Villejuif, France; 4) Advanced Biomedical Computing Center, SAIC-Frederick, NCI, Frederick, MD, USA; 5) Center for Biomedical Informatics, Advanced Biomedical Computing Center, SAIC-Frederick, NCI, Frederick, MD, USA; 6) 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, California; 2) Deps of Biological Chemistry and Computer Science, UCLA, Los Angeles, California.

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

495F Spread of repressive chromatin from the expanded GAA trinucleotide repeat mutation contributes to gene silencing in Friedreich ataxia. Y.K. Chudakov1, S.J. Bidichandani1,2, 1) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Department of Biochemistry & Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

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

496T Complex molecular findings in 11p15-associated Imprinting Disorders require a broadening of testing strategies. M. Begemann, S. Spengler, L. Soellner, T. Eggemeer. RWTH Aachen, Aachen, Germany.

The chromosomal region 11p15 is one key player in molecular processes regulated by genomic imprinting. Genomic as well as epigenetic disturbances affecting the two imprinting control regions (ICRs) 11p15 are associated either with Silver-Russell syndrome (SRS) or Beckwith-Wiedemann syndrome (BWS): SRS patients carry ICR1 hypomethylation, maternal 11p15 duplications or uniparental disomies (UPD), whereas the opposite alterations are characteristic for BWS. In the last years, a growing number of patients suffering from imprinting disorders (IDs) has been reported showing a hypomethylation at the two ICRs at 11p15 as well as at further imprinted loci on other chromosomes. The molecular basis of these multilocus methylation defects (MLMDs) is widely unknown, however an interaction between trans-localized imprinted genes via a so-called Imprinted Gene Network (IGN) has been suggested. Cases with other types of molecular disturbances (e.g. imprinting defects (MLMDs) and chromosome aberrations) support this IGN concept. The complex molecular alterations as well as the overlapping and sometimes ambiguous clinical findings in ID patients often make the decision for a specific ID test difficult. As aforementioned, 11p15 loci are consistently affected in all ID patients with more complex alterations, but the pattern of affected loci is nearly unpredictable.

We therefore suggest to implement molecular tests in routine ID diagnostics which allow the detection of a broad range of (epi)mutation types (epimutations, UPDs, chromosomal imbalances) and cover the currently known ID loci. The need to apply multilocus tests (methylationspecific MLPA, SNUPE) is corroborated by our experience from routine diagnostics in more than 650 patients referred as SRS or BWS: (a) multilocus tests increase the chance to identify even slight mosaic hypomethylation patterns is growing. (b) Patients with unusual phenotypes and unexpected molecular alterations will be detected. (c) The testing of rare imprinting disorders becomes more efficient and quality of molecular diagnosis increases. (d) The new cases will help us to decipher the complex regulation of ACR. We therefore suggest to implement molecular tests in routine ID diagnostics.
497
Allelespecific analysis of putative imprinted gene network members in human. D. Prawit1, F. Bohne1, D. Langer1, U. Martini1, K. Oexle1, B.U. Zabel1, T. Enklaar1, 1) Molecular Genetics Lab. Center for Pediatrics and Adolescent Medicine, Univ. Medical Center, Mainz, Germany; 2) Institute of Human Genetics, Technical University, Munich, Germany; 3) Center for Pediatrics and Adolescent Medicine, University Medical Center, Freiburg, Germany.

The two imprinting disorders Beckwith-Wiedemann syndrome (BWS) and Silver-Russell syndrome (SRS) present with marked intrauterine and postnatal overgrowth (BWS) or growth retardation (SRS). Molecular defects are heterogeneously associated and are mainly due to epigenetic changes in imprinting control regions (ICR) of the 11p15.5 region. ICR1 regulates the monoallelic expression of Igf2 and H19 in cis. A subgroup of SRS and BWS patients present with epigenetic alterations of other chromosomal regions, arguing for a functional dependence of the affected imprinted genes. To date the molecular mechanisms of the ICR1 effects are incompletely defined and their analysis often has to be performed using genetically engineered model organisms. The description of a network of imprinted genes (IG) in mice raised the possibility to improve the comprehension of multilocus imprinting defects in humans and depicts candidate genes involved in associated clinical symptoms. The presented work offers an unusual mechanistic insight into ICR1-driven regulation of imprinted genes in human cells. By utilizing unique fibroblasts with paternal (BWS) or maternal (SRS) uniparental disomy 11p15, we uncovered a set of transcriptionally co-regulated imprinted genes on different chromosomes. According to findings for murine orthologs the genes can be part of an IG involved in regulating developmental growth.

498T
The allelic expression of Gnas is affected by mutations in Mecp2 and maternal strain. H. Stilk1,2, R.D. O’Connor1,2, A. Kumar1, N.C. Schanen1,2.
1) Department of Biological Sciences, University of Delaware, Newark, DE 19716, USA; 2) Nemours Biomedical Research, Alfred I duPont Hospital for Children, Wilmington, DE 19803, USA.

The GNAS gene encodes the G protein stimulatory alpha subunit of (Gs) and shows tissue specific monoallelic expression that is regulated through a methylation-dependent process. To determine whether the Methyl CpG Binding Protein 2 (Mecp2) was involved in regulation of imprinting of the Gnas transcript, we examined allelic expression of Gnas in tissues from male progeny of an outcross between C57BL/6J females, wild type or heterozygous for a null allele for Mecp2 (Mecp2+/-Bird/J), and Mus castaneus (CAST) males. Using a strain specific polymorphism, we determined the parental origin of the Gnas transcript in tissues at postnatal days 3 (P3) and 28 (P28). In keeping with previous reports, our control cross showed a preference for maternal expression of Gnas in tissues that are imprinted. However, unexpectedly, we found that allelic expression of Gnas was altered in both wildtype and mutant progeny of the Mecp2+/- mothers, showing preferential paternal expression of the Gnas transcript in a number of tissues. Furthermore, we observed both a loss as well as a gain of imprinting at P3 in the kidney and calvaria, respectively. Since a strain effect has been reported in other imprinted loci, we sought to determine whether the strains used in this outcross could be affecting the expression of Gnas by performing a reciprocal cross between CAST female X C57BL/6J male mice. These studies revealed expression patterns consistent with the previous wildtype outcross with the exception of the spleen, which showed preferential expression of the C57BL/6J allele. Additionally, we examined the differentially methylated region (DMR) in exon 1A that is responsible for controlling expression of Gnas using sodium bisulfite conversion and sequencing analysis to determine if there was a developmental shift in methylation in wild type C57BL/6J mice. In all tissues examined, methylation increased significantly from P3 to P28. These studies suggest a transgenerational effect of maternal Mecp2 genotype on imprinting of Gnas in mice that occurs in a tissue specific manner and demonstrate epigenetic regulatory processes are impacted by strain and age.

499F
Exploring the role of NLRP7 in reprogramming of CpG methylation in the oocyte and early embryo. S. Mahadevan1,2, S. Wan1,2, Y. Wool1,2, Z. Zwell1,2, V. Van den Berghe1,2.
1) Department of Translational Biology and Molecular Medicine, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics and Gynecology, Baylor College of Medicine, Houston, TX; 3) Jan and Dan Duncan Neurological Research Institute at Texas Children’s Hospital, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics (Neurology), Baylor College of Medicine, Houston, TX; 6) Lillehei Heart Institute and Department of Pathology, University of Minnesota, Minneapolis, MN; 7) Department of Pathology and Laboratory Medicine, McMaster University, Hamilton, ON, Canada.

Maternal effect mutations in NLRP7 cause recurrent, biparental hydatidiform mole (BHIM), a trophectodermal disease by loss of maternal imprinting at maternally imprinted loci. The existing paradigm suggests a loss of methylation at maternally imprinted loci, while studies from our group in hESC revealed a global, genome wide shift in methylation dynamics causing us to further investigate the extent to which loss of NLRP7 affects the cellular epigenetic state. We had previously shown that NLRP7 physically interacts with the transcription regulator YY1. ChiP-qPCR of YY1 at imprinted loci such as PEG3 and SNRPN. This increased occupancy of YY1 at its target sites is not caused by an upregulation in YY1 expression or intra-cellular redistribution within the cell. Fortuitously, it was also observed that NLRP7, which was previously thought to be entirely cytoplasmic in localization, was capable of migration into the nucleus. Given that YY1 is also an important regulator in the nucleus, NLRP7’s presence in the nucleus likely contributes to the observed differences in YY1 occupancy at its target sites. Computational interrogation of the protein sequence of NLRP7 revealed a putative bipartite nuclear localization signal (NLS). In the absence of NLRP7’s ability to localize within the nucleus, it will be informative to assess whether YY1 occupancy at its target sites is also affected. This information is not only valuable from the perspective of understanding NLRP7 function but is suggestive of NLRP7’s potential of being a novel YY1 cofactor. The observation that the loss of methylation at maternally imprinted loci is caused by a maternal loss of NLRP7 argues in favor of a role for this protein in the oocyte and early embryo in the process of imprint acquisition and/or maintenance. A systematic interrogation of proteins with known roles in imprint acquisition and/or maintenance revealed physical interaction of NLRP7 with CG Binding Protein 1 (CGBP / CF1). Given that distinct domains of CFP1 function as mediators of crosstalk between histone and DNA methylation, current efforts are focused on delineating which domains of CFP1 interact with NLRP7. The overall scope of this project is to gain insight into the roles of these proteins in the reprogramming of CpG methylation in the oocyte to embryo transition.

500T
Epigenetic effects of the endocrine disruptor phthalate: influence of the genetic background. A. Paolioli-Giacobino1,2, Ch. Stouder1, E. Somm2.
1) Department of Genetic Medicine and Development, Geneva University Medical School, Geneva, Switzerland; 2) Swiss Center for Applied Human Genetics, Geneva University Hospital, Geneva, Switzerland.

Di-(2-ethylhexyl)phthalate (DEHP) is a plasticizer with endocrine disrupting properties found ubiquitously in the environment. The aim of the present study was to evaluate the effects of DEHP exposure in the first trimester of pregnancy in pregnant mice, on imprinted genes in the offspring. Mice of 2 different genetic backgrounds (C57BL/6 and FVB/N) were chosen and treated with DEHP during the time of embryo sex determination. Potential effects of DEHP on the methylation of the differentially methylated domains (DMDs) of 4 paternally expressed genes, Smnp, Peg1, Peg3 and Ifg2r and of 3 maternally expressed genes, H19, Gtl and Ifg2 were investigated. The degrees of methylation of the imprinted genes were analysed in the sperm as well as in the liver and skeletal muscle. In parallel, the mRNA expressions of Smnp, Peg3, Ifg2r in these same tissues were studied to examine the link between the imprinted gene methylation and expression. In the sperm DEHP did not affect the degree of methylation of the paternally expressed genes except that of Ifg2r in the C57BL/6 mice but induced decreases in the degree of methylation of all the maternally expressed genes in the C57BL/6 mice whereas it had no effect on these same genes in the FVB/N mice. In the liver, DEHP induced ubiquitous effects, potentially representing a relaxation of the monoallelic methylation pattern, in all in 6 out of 7 of the imprinted genes in the C57BL/6 mice, respectively, in the skeletal muscle, 5 out of 7 and 3 out of 7 imprinted genes were affected by DEHP in the C57BL/6 and FVB/N mice, respectively. Eventually, the mRNA expressions in the sperm, liver and muscle of 2 imprinted genes, Smnp and Peg3 were increased by DEHP in FVB/N mice, in the absence of any change of DMD methylations. In conclusion, DEHP has an effect on sperm DNA methylation that seems to be dependent on the genetic background. The direct effects of DEHP on somatic cells affect practically all imprinted genes studied, especially in the liver. The overall consequences of DEHP on imprinted genes were either direct or mediated by epigenetic effects other than DNA methylation.
501F Establishing induced pluripotent stem cell-derived neurons as an appropriate cellular model for studying mechanisms of myoclonus-dystonia. K. Freimann, A. Westenberger, P. Seibler, A. Weissbach, N. Brueggemann, K. Lohmann, C. Klein, A. Grunenwald. Institute of Neurogenetics, University of Luebeck, Luebeck, Germany.

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

502T Full resolution DNA methylethanol analysis in multiple tissues from twins. S. Busche1,2, M. Caron2, T. Kwan3, V. Forgetta4,5, B. Ge2, S. Westfall6, J. B.T. Bell7, H. von Mutters2, R. B. Carroto2,4,5, J. T. Ball4,5, J. G. W. T. Roger8,9, J. H. T. F. B. H. White8,9,10, J. B. Richards1,4,5, G. Bourque1,2, J. M. Lathrop1,2, P. Deloukas4,10, A. D. Macarthur1,2, T. Pastinen1,2, E. Grundberg1,2.

1. Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 3) Canada; 4) Department of Twin Research and Genetic Epidemiology, King’s College London, London, United Kingdom; 5) Genome Trust, Sanger Institute, Hinxton, Cambridge; United Kingdom;

Numerous population studies applying methylethanol analyses by the Illumina450K array are on-going, but no comprehensive or unbiased analysis of CpG variation at the population level has been carried out to date. Using adipose tissue (AT) and whole-blood (WB), from monzygotic (MZ) and dizygotic (DZ) twins of the well characterized MuTHER/TwinsUK cohort (e.g. dense phenotype, 450K and WGS data available) we aimed to unravel the impact of stochastic, environmental, and genetic factors underlying methylethanol (mCpG) variation and the impact on disease susceptibility using whole genome bisulfite sequencing (WGBS). So far, we have analyzed 26 AT and 21 WB samples (6 MZ pairs, 5 DZ pairs, 7 singlets) generating a total of 1.97 Tbp sequence at an average 6.5-fold coverage for AT and 9.7-fold for WB samples. Totally, we identified 21M and 24M CpG sites in AT and WB with an average methylation level of 85%, CpG values detected by WGBS at > 4-fold coverage were highly concordant with overlapping Illumina 450K data (Pearson’s r = 0.93-0.95) but only <1% of global mCpG variation is accessible through the 450K array. pairwise comparisons of MZs (<20% vs. >80% methylated) allowed us to identify the proportion of mCpG variation that is of environmental origin, which corresponded to on average 0.03% in AT and 0.015% in WB. Similar comparisons of DZs and unrelated yielded a ~5- and ~10-fold increase. Sequence variants at the CpG site appeared to underlie most non-environmental mCpG, and thorough variant removal reduced this increase to ~2-fold. Further analysis of environmental driven differentially methylated regions (eDMRs; >3 consecutive CpGs within 500 bp and >5mCpG>40%) using MZs revealed on average 90 eDMRs per tissue and pair with no overlap across tissue, highlighting tissue-specificity with likely correlation with phenotypic discordance. We also studied tissue-specific DMRs (tDMRs) and found AT-specific, hypomethylated tDMRs to be highly enriched (17-fold) in human adipocyte enhancer elements (HSAKm4m1, NIH Roadmap Epigenomics Project). Finally, the existence of WGS (UK10K project) will allow us to define the largest allele-specific methylation (ASM) dataset in multiple tissues, including random, parent-of-origin and sequence-dependent effects. Preliminary data indicates ASM effects accounting for 1% of all CpGs, where a minor proportion (3.4%) corresponds to known imprinted loci with the remaining effect most likely being due to sequence-dependent ASM effects.

503F Epigenome-wide association study on identical twins discordant for birth weight. Q. Tan1,2, M. Frost1, L. Christiansen1.

1) University of Southern Denmark, Odense, Denmark; 2) Odense University Hospital.

Epidemiological evidences have shown that early life exposure to adverse environment, e.g. bad nutrition, and stress during pregnancy of mother which usually resulting in low birth weight, could have long-term health consequences including metabolic disorders, diabetes, hypertension, and even mental diseases in adult life. Epigenetics is a molecular mechanism that explains the acquired changes or modification in gene function independent of DNA sequence variation with DNA methylation as one of several epigenetic mechanisms most intensely studied. In order to examine if low birth weight induces persistent epigenetic modification detectable at adult ages, we performed a genome-wide DNA methylation profiling in peripheral blood of 150 pairs of identical Danish twins discordant for birth weight using the Illumina Infinium HumanMethylation450 BeadChip featuring 485,000 CpG sites across the genome. After quality control and data preprocessing using free R-package minfi, data were analysed by a mixed effects model including mixed fixed effect variables such as birth weight such as birth weight, sex, age and sex of twin pairs; random effect variables such as batch, well, and sample position on the array, etc. Statistical analysis revealed 12 probes with p value<1e-05, among them 1 probe with p value<1e-06. Biological pathway analysis using these probes showed no significant functional cluster involved. In conclusion, our analysis suggests that discordant birth weight in identical twins is not associated with measured DNA methylation level at adult ages. Similar studies on tissues other than peripheral blood are required in order to re-confirm and generalize our conclusion.
504T Identification of a differentially methylated locus in monozygotic twins discordant for esophageal atresia. L. Boutaud de la Combe1, A. Pelet1, C. Boile2, J. Tost1, W. Carpenter3, N. Cagnard2, R. Smigiel1, V. Gaudin2, J. Amiel1, S. Lyonnet1. 1) INSERM U781, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, Paris 15, Paris, France; 2) Plateforme de génomique et de bioinformatique, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, Paris 15, Paris, France; 3) Centre National de Génotypage CEA - Institut de Génomique 2 rue Gaston Crémieux CP5721 91057 Evry Cedex France; 4) Plateforme P3S,UPMC Pitié Salpêtrière 91 Blvd de l'Hôpital 75013 Paris France; 5) Genetics Depart- ment Wroclaw Medical University Marcinkowskiego 1 PL 50-383 Wroclaw, Poland; 6) Service de Chirurgie Viscérale Pédiatrique Hôtel Necker 149 rue de Sèvres 75743 Paris Cedex 15, France; 7) AFAO 56 rue Cécile 94700 Maisons-Alfort France.

Until recently monozygotic twins were considered to be genetically identical. However the increasing description of monozygotic twins discordant for a phenotype leads us to change our point of view. Numerous causes can explain the discordance between twins. In addition to environmental causes, genomic and epigenomic modifications can occur in only one of the sibs, such as mosaic mutations, post-zygotic alterations and DNA methylation. In order to understand the molecular bases of isolated esophageal atresia, we are studying monozygotic twins discordant for this serious and frequent congenital malformation. Our work aims at finding the origin of the esopha- geal atresia that occurs in only one child of each twin pair. We decided to investigate two hypotheses: genomic and epigenomic. Eight twin pairs were studied; one twin of each pair has esophageal atresia with or without other congenital abnormalities (such as anomalies of the heart or kidney). All the pregnancies were based on natural conception. The twins are consanguineous. Twins were raised together and the children are 2 to 7 years of age. Thus, we assume that environmental differences are negligible. Genomic and epigenomic hypothesis were tested searching first for copy number variations (CNVs) in the twins DNA and then for methylation differ- ences between the child suffering from esophageal atresia and his healthy co-twin. No CNVs were found, however, using a cytoseine microarray (Illum- ina Infinium HumanMethylation450 BeadChip) we have compared 450,000 single nucleotide polymorphism (SNP) and DNA methylation in monozygotic twins discordant for esophageal atresia that occurs in only one child of each twin pair. The phenotype value closest to the DNA extraction date was choosen. The phenotype value closest to the DNA extraction date was choosen. The total forearm, hip and spine BMD. The methylation difference of a total of 33 pairs of MZ twins, 23 pairs had phenotype values for total forearm, hip and spine BMD. The methylation difference of a total of 7 genes had significant Spearman correlation (p < 1x10-4) with differences in BMD phenotypes; 3 for TH (NTRK3, VSTM1 and 2NFP438), 3 for FA (PIGC, PKP1, CBG), and 2 for SP (CBG and SLC6A11). The most significant correlation was obtained for corticosteroid-binding globulin (CBG) using the total forearm BMD phenotype (p-value 5.46x10-6), and was also significant, albeit to a lesser extent, for total spine BMD (p-value 1.67x10-5). The correlation for CBG is highly linear and positively correlated, where a positive difference in methylation leads to a positive difference in BMD phenotype between MZ sib-pairs. Discussion: This epigenome wide study has identified 7 putative loci correlated with BMD phenotypes. Among these loci is corticosteroid-binding globulin, a gene known to influence corticosteroid metabolism which in turn strongly influence bone physiology. Further validation is underway shortly, including replication of findings using a larger cohort of twins and unrelated individuals using the Illumina 450k methyla- tion kit.

505T Analysis of the epigenetic interplay between DNA methylation, histone modifications and gene expression in monozygotic twins discordant for psoriasis identified dysregulated disease-associated genes. R. Lyle1, K. Gervin1,2, G.D. Gillifan1,2, M. Hammers1, H.S. Hjorthaug1, A.O. Olsen1,2. 1) T. Hughes1, J.R. Harris2, D.E. Undlien1,2. 1) Med Gen, Oslo Univ Hosp Ullevål, Oslo, Norway; 2) University of Oslo, Oslo, Norway; 3) Depart- ment of Dermatology, Oslo University Hospital, Oslo, Norway; 4) Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway.

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

505F Genome wide DNA methylation profiling of monozygotic twins discord- ant for trisomy 21. M.R. SAILAINI1,2, F. Santoni1, A. Letourneau1,2, P. Makrythanasis1, C. Borel1, M. Guipponi1, C. Gehrig1, A. Vannier1, S.E Antonarakis1,2,3. 1) Genetic MEDICINE and DEVELOPMENT, University of Geneva, Geneva, Geneva, Switzerland; 2) National Center of Competence in Research Frontiers in Genetics, University of Geneva, Switzerland; 3) iGE3 institute of Genomics and Genomics of Geneva, Switzerland.

DNA methylation is essential in mammalian development and has an effect on gene expression. We have hypothesized that methylation differences induced by trisomy 21 (T21) contribute to the phenotypic characteristics and heterogeneity in T21. In order to determine the methylation differences in T21 without the interference of the genomic variation, we have used fetal fibroblasts from monozygotic twins discordant for T21, normal monozygotic twins without T21, and unrelated normal and T21 individuals as controls. We applied Reduced Representation Bisulftite Sequencing (RRBS) to generate nucleotide resolution of DNA methylation based on high throughput sequenc- ing (HiSeq 2X100bp, One sample per lane) between each pair of twins. CpGs with at least 10X read coverage in two technical replicates were sel- ected for the subsequent analysis. An initial analysis of MZ twins discor- dont for T21 identified 28,526 differentially methylated C nucleotides out of 1,589,507 interrogated nucleotides (DMCs) (FDR<0.001 and at least 50% methylation difference per C). The KEGG gene ontology analysis of genes harboring these DMCs, showed an enrichment for calcium signaling (FDR 3.1x10-7), axon guidance (FDR 7.4x10-7), and focal adhesion (FDR 5.1x10-7) pathways. This preliminary study of methylation differences in monozy- gotic twins discordant for genomic abnormalities is a promising approach to understand the molecular pathophysiology of aneuploidies.
508T DNA methylation analysis of the human X chromosome in multiple tissues. A. Cotton 1,2, M. Jones 1,3, C. Chen 1,4, W. Wasserman 1,2, M. Kobor 1,2, C. Brown 1,2. 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Molecular Epigenetics Group, Life Sciences Institute, Vancouver, BC, Canada; 3) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; 4) Graduate Program in Bioinformatics, University of British Columbia, Vancouver, BC, Canada.

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

509F Toward understanding the higher-order chromatin organization at the human inactive X chromosome. E. Darrow, B. Chadwick. Biological Science, Florida State University, Tallahassee, FL.

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

510T The p.V37I Exclusive Genotype Of GJB2: A Genetic Risk-Indicator of Postnatal Permanent Childhood Hearing Impairment. L. Liu 1,2, Z. Tao 1,2, Q. Huang 1,2, Y. Chai 1,2, X. Li 1,2, Z. Huang 1,2, Y. Li 1,2, M. Xiang 1,2, J. Yang 1,2, G. Yao 1,2, Y. Wang 1,2, T. Yang 1,2, H. Wu 1,2. 1) Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; 2) Ear Institute, Shanghai Jiaotong University, Shanghai, China; 3) Shanghai Children’s Medical Center, Shanghai, China; 4) Shanghai Child Health Care Institute, Shanghai, China.

Background: Postnatal permanent childhood hearing impairment (PCHI) is frequent (0.25%–0.99%) and difficult to detect in the early stage, which may impede the speech, language and cognitive development of affected children. Genetic tests of common variants associated with postnatal PCHI in newborns may provide an efficient way to identify those at risk. We hypothesized that the p.V37I exclusive genotype of GJB2, a prevalent genetic variant in East Asians, is associated with postnatal PCHI and therefore serves as a genetic risk-indicator. Methods and Findings: We sequenced the GJB2 gene in a cohort of 45 Chinese subjects with postnatal PCHI, 1516 ethnically-matched control newborns with various newborn hearing screen (NHS) results, and 173 newborns who failed the NHS but passed the referral hearing diagnosis. The p.V37I exclusive genotype of GJB2 is present in a substantial percentage (20%) and is strongly associated with postnatal PCHI (P = 1.4x10−10). The prevalence of this genotype also increases significantly in normal-hearing newborns who failed at least one NHS (P = 0.024 for those who failed only the initial NHS; P = 1.7x10−8 for those who failed both the initial and the repeat NHS). Conclusion: The p.V37I exclusive genotype of GJB2 causes subclinical hearing impairment at birth and increases risk for postnatal PCHI. Genetic testing of GJB2 in East Asian newborns may provide an efficient way to identify those at risk. In addition to importance in modulating disease, the study of XCI provides new insights into epigenetics gene regulation and nuclear compartmentalization.

511F Up-Regulates of angiogenesis-associated MicroRNAs in placenta from women with a low flow-mediated vasodilation. L.M. RODRIGUEZ, A.F. DUQUE, A. SANCHEZ, A.C. AGUILAR, A. CASTILLO. UNIVERSIDAD DEL VALLE, CALLE 4B 36-00, CALI - COLOMBIA.

Purpose: microRNAs profile have emerged as a molecular tool to understand epigenetic regulation process and a possible impact in diseases progression or complications, as in future mother at gestational period, during which endothelial function changes may be involved in differential expression profile of microRNAs associated with angiogenesis in placental vascularization. Methods: we evaluated FMD (flow-mediated vasodilation) during the first and third trimester of pregnancy in eleven primigravidae singletons as a vascular functional biomarker. Besides, at the delivery we took a placental sample to evaluate the microRNA expression associated with angiogenic pathway using miScript miRNA PCR array (Qiagen). Mann-Whitney and t-Student test, with p-value <0.05 were use to statistical analysis. Results: we found four upregulated miRNAs: hsa-miR-16 (p-value: 0.044941), hsa-miR-130a (p-value: 0.037793), hsa-miR-132 (p-value: 0.04386) and hsa-miR-296 (p-value: 0.049006) at women group that showed a negative change in FMD. Three of them (hsa-miR-130a, 132 and 296) are pro-angiogenic with a possible implication in a placental blood vessel increasing production and a decrease blood flow compensatory mechanism. Conclusion: Thereby, the impact in the microRNAs identification in the maternal-fetal health is important to prevent possible problem associated to placental vascularization.
512W
ChIP-seq ascertained occupancy of MEF2C, a GWAS-implicated osteoporosis locus, points to an inflammation-mediated role in bone density determination. S.F. A. Grant1, S. Delard2, F. Zhu2, O. Xia1, A.D. Woods2, K.D. Hankenson3, M.E. Johnson4. 1) Division of Human Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA; 4) Pathology and Laboratory Medicine, The Children’s Hospital of Philadelphia, Philadelphia, PA.

Genome-wide association studies (GWAS) have demonstrated that genetic variants regulate the binding of MEF2C (MADS box transcription enhancer factor 2, polyepitope C) locus is robustly associated with bone mineral density (BMD), primarily at the femoral neck. MEF2C is a transcription factor known to operate via the Wnt signaling pathway. Our hypothesis was that MEF2C regulates the expression of a set of molecular pathways critical to skeletal function. Drawing on our laboratory and bioinformatic experience with ChiP-seq, we analyzed ChIP-seq data for MEF2C available via the ENCODE project to gain insight into its global genomic binding pattern. We aligned the ChIP-seq data generated for GM12878 (an established lymphoblastoid cell line). Using HOMER, a total of 17,611 binding sites corresponding to 8,118 known genes were observed. We then performed a pathway analysis of the gene list using Ingenuity. At 5kb, the gene list yielded ‘EIF2 Signaling’ as the most significant annotation, with a P-value of 5.01 x 10^-38. Moving further out, this category remained the top pathway at 50kb and 100kb, then dropped to just second place at 500kb and beyond by ‘Molecular Mechanisms of Cancer’. Also consistently high on these lists at all distances was ‘B Cell Receptor Signaling’ and ‘Cell Cycle Receptor Signaling’, plus at 50kb and beyond ‘Cell Adhesion’ in concert with the inflammation signature and resonates with the main general finding from GWAS of bone density.

We also observed that MEF2C binding sites were significantly enriched primarily near inflammation associated genes identified from GWAS; indeed, a similar enrichment for inflammation genes has been reported previously. Using a similar approach for the vitamin D receptor, an established key regulator of bone turnover. These results represent the first ever reported ChIP-seq derived genome wide map of MEF2C binding. Our analyses point to known connective tissue and skeletal promoters but also provide insights into networks involved in skeletal regulation. The fact that a specific GWAS category is enriched points to a possible role of inflammation through which it impacts BMD.

513T

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

514F
Human-specific cytogenetic structures support the emergence of new regulatory elements. G. Giannuzzi, A. Reymond. Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland.

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

515W
Architectural proteins modulate the higher order chromatin structure of the CFTFR locus. N. Gosalia1,2, A. Harris1,2. 1) Human Molecular Genetics Program, Ann and Robert H. Lurie Children’s Hospital of Chicago Research Center, Chicago, IL; 2) Department of Pediatrics, Northwestern University, Feinberg School of Medicine, Chicago, IL.

Cystic fibrosis is an autosomal, recessive disease that results from mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, which encodes a chloride ion channel with a complex expression pattern. Tissue-specific CFTR expression is regulated by multiple cis-acting elements, including intronic enhancers, which interact with the promoter. The mechanisms that organize higher order chromatin structure to establish and maintain gene expression are not fully understood; however, data suggest they may involve CTCF and cohesin. These factors have important roles in the three-dimensional organization of loci and at insulators, which are critical barriers for preventing inappropriate activation or repression of genes. At the CFTR locus, the CTCF and cohesin components, RAD21, increase gene expression. CHD for CTCF and the cohesin components, RAD21 and SMCR1, after CTCF or RAD21 knockdown shows loss of all three factors at several sites across the region and retained occupancy of RAD21, increases gene expression. CHI for CTCF and the cohesin components, RAD21 and SMCR1, after CTCF or RAD21 knockdown shows loss of all three factors at several sites across the region and retained occupancy of RAD21. To determine if loss of CTCF and/or RAD21 altered the higher order organization of CFTR, quantitative chromosome conformation capture (q3C) was used. After RAD21 depletion, q3C data show partial loss of interactions across the locus between the gene promoter, known enhancers such as the one in intron 11, and CTCF/cohesin binding sites. In contrast, after CTCF knockdown no changes in enhancer-promoter interactions are observed, however, interactions between the -20kb insulator element and CTCF/cohesin binding elements 3’ to CTR are suppressed. These data suggest that CTCF and the cohesin complexes are critical for some aspects of the chromatin organization at the locus they are not the major components of the complex that loops distal enhancers to the promoter. Moreover, the increase in CR expression after knockdown also suggests that CTCF and cohesin may inhibit CR expression by recruiting repressive complexes to the locus. Identifying these response factors may facilitate the therapies to alleviate cystic fibrosis disease severity by increasing the levels of CR transcript and CR protein.
Impact of three-dimensional organization of chromatin on long-range enhancers. A. Pankov, J. Song.

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

Genetic determinants of population-level variation of chromatin modifications. H. Klippen1,2, S.M. Waszak3, A.R. Gschwind1,2, S.K. RagHAV2,3, R.M. Witwicki4, A. Otroi4, M. Wiederkehr4, N. Panousis2,3, N. Hernandez1,2, A. Raymond1,2, B. DePlancke1, E.T. Dermitzakis1,2,1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Institute of Bioengineering, School of Life Sciences, Swiss Federal Institute of Technology, Lausanne, Switzerland; 4) Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland.

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

Molecular analyses of cis-interactions of the SOX9 promoter with its sex determining long-range regulators using chromatin conformation capture-on-chip (4C). P. Stankiewicz1,2, M. Smyk2, P. Szafarski1,2,1) Dept. of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept. of Medical Genetics, Institute of Mother and Child, Warsaw, Poland. Evolutionary conserved transcription factor SOX9 is essential for development of testes and differentiation of choroid plexus. Heterozygous point mutations and genomic deletions involving SOX9 lead to the skeletal malformation campomelic dysplasia (CD) often associated with male-to-female sex reversal. Chromosomal rearrangements with breakpoints mapping in the protein-coding desert regions up to 1.3 Mb both 5' and 3' to SOX9 and likely disrupting cis-regulatory elements, have been described in patients with usually milder forms of CD. Breakpoint clustering 5' to SOX9 allowed to define four non-overlapping intervals associated with different phenotypes: Pierre Robin sequence (~1.06-1.23 Mb), acampomelic CD (~789-932 kb), sex reversal (~517-595 kb), and moderate to severe CD (~50-375 kb). In addition, a different phenotype consistent with Cooks syndrome (brachydactyly-anonychia) has been reported as associated with genomic duplications (~1.2 Mb 5' to SOX9. We found that these regions overlap four gene clusters encoding long non-coding RNAs (IncRNAs): TCONS_00025479, TCONS_00025148, TCONS_00025195, and TCONS_00026251, suggesting that similar to other developmental genes, incRNAs might contribute to long-range regulation of SOX9 expression. The sex reversal region (RevSex) was proposed to harbor a putative testis-specific and sex determining enhancer. The other sex determining interval was mapped to a gene desert >1.3 Mb downstream to SOX9. We performed chromatin conformation capture-on-chip (4C) analysis in Sertoli cells and lymphoblasts to verify the already proposed long-range interactions and to identify potential novel regulatory elements responsible for sex reversal in patients with CD. We identified several novel cis-interacting regions both up- and downstream to SOX9 with some of them overlapping lncRNAs preferentially expressed in testes. Our data further support the role of two previously identified regulatory regions in the control of SOX9 expression during sex determination and point at lncRNAs as likely mediators of some of these interactions.

Identification of genetic variants that affect histone modifications in human cells. G. Mcvicker1,2, B. van de Geijn1,2, J.F. Degner3,4, E.C. Cain3, N.E. Banovich1, N. Lewellen2, M. Mythili2, Y. Gilad2, J.K. Pritchard3,2, 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Howard Hughes Medical Institute, University of Chicago, Chicago, IL; 3) Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, IL.

Histone modifications are important markers of function and chromatin state, yet the DNA elements that direct them to specific locations in the genome are poorly understood. Here we use the genetic variation in Yoruba lymphoblastoid cell lines as a natural experiment to identify genetic differences that affect histone marks and to better understand their relationship with transcriptional regulation. Across the genome, we identified hundreds of quantitative trait loci that impact histone modification or RNA polymerase (PolII) occupancy. In many cases the same variant is associated with quantitative changes in multiple histone marks and PolII, as well as in DNase sensitivity and nucleosome positioning, indicating that these molecular phenotypes often share a single underlying genetic cause. We find that polymorphisms in many transcription factor binding sites cause differences in local histone modification and identify specific transcription factors whose binding leads to histone modification in lymphoblastoid cells. Finally, we find that variants that impact chromatin at distal regulatory sites frequently also direct changes in chromatin and gene expression at associated promoters. In summary, the class of variants identified here generate coordinated changes in chromatin state locally and sometimes at distant locations, frequently drive changes in gene expression, and likely play an important role in the genetics of complex traits.
Identification of copy number variants (CNVs) using computational algorithms and array CGH technologies. P. S. Samarakoon1,2, H. S. Sorte1,2, B. E. Kristiansen1,2, T. Skodje2, A. Stray-Pedersen2, O. K. Rødningen1,2, R. Lyle1,2, H. S. Sorte1,2, B. E. Kristiansen1,2, T. Skodje2, A. Stray-Pedersen2, O. K. Rødningen1,2, R. Lyle1,2

1 Department of Medical Genetics, University of Oslo, Oslo, Norway; 2 Department of Medical Genetics, Oslo University Hospital, Oslo, Norway.

With advances in next generation sequencing technologies and genomic capture techniques, exome sequencing has become a cost-effective approach for mutation detection in genetic diseases. However, computational CNV prediction algorithms using exome sequence data exhibit limitations due to the low sensitivity and specificity in predicting small CNVs (covering 1-4 exons). Additionally, Comparative Genomic Hybridization (CGH) microarrays used for genome-wide high-resolution CNV detection also show restrictions in exonic CNV detection due to the low scoring probes over GC rich regions. Therefore, the goal of the research project was to develop a protocol to detect exonic CNVs (including 1-4 exons), combining computational algorithms and custom aCGH. In this study, we predicted CNVs of 30 exomes obtained from 1000 genomes project using seven computational programs, including an in-house developed algorithm. CNV calling of the proposed algorithm relies on GC% based read depth normalization followed by generation of the best-ranked reference dataset for target exome. In parallel to the computational prediction, CNVs were also identified by using aCGH, which was designed to capture exonic regions in 1000 genomes exome and GenCode v.15. Next, true CNVs were identified by comparing computational predictions to the aCGH results. Results of each computational program showed a significant variation in predicted CNV counts (ranging 0-2500). Due to the drastic variation in resulted CNV counts, intersection operations were performed to identify CNVs predicted by multiple programs. As intersection operation showed a direct effect in decreasing the number of predicted CNVs to 0-250 from 0-2500, this was used as a filter for our CNV pipeline. Finally, CNVs resulted from complete protocol (true CNVs), which were not reported in database of genomic variants (DGV) were identified as novel CNVs. Application of the complete protocol on 9 individuals presented 77 novel CNVs including 26 that cover single exons. Additionally, protocol was then implemented on complete datasets (true CNVs), which were not reported in database of genomic variants (DGV). From these results, we concluded that a combination of computational and custom aCGH methods is essential to identify novel CNVs and PIDD causing CNVs from the computational approach were validated using the custom array. Thus, identifying PIDD causing CNVs while demonstrating the capability in detecting shorter exonic CNVs proved clinical importance of the proposed protocol.
523F
Genomic characterization and copy number analysis of human induced pluripotent stem cells (iPSCs) from 22q11.2 deletion syndrome patients. C. Purmann1,2, S. Pasca1, H. Gai1, A. Krawisz1, X. Zhu1,2, J. Rapoport4, J. Berstein5, J. Hallmayer1, R. Dolmetsch1, A.E. Urban1,2, 1) Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Neurobiology, Stanford University, Stanford, CA; 4) National Institutes of Health, Child Psychiatry Branch, NIMH, Bethesda, MD; 5) Department of Pediatrics, Stanford University School of Medicine, Stanford, CA.

The 22q11.2 deletion syndrome, also known as DiGeorge Syndrome or Velocardiofacial Syndrome (VCFS), is a common genomic disorder in humans. The clinical phenotype is variable, but includes notably neurodevelopmental abnormalities, craniofacial and cardiovascular malformations, as well as immune deficiencies. It is thought that large CNVs present critical aberrations during early brain development that manifest in childhood or later in life. The lack of understanding how exactly these CNVs exert their influence on a molecular level is exacerbated in neurogenetic diseases where it is difficult to access the relevant human tissues for detailed molecular studies. One very promising avenue around this obstacle comes from the recent development of induced pluripotent stem cells (iPSCs) whereby skin cells from patients can be reprogrammed into pluripotent stem cells which can then be differentiated into neuronal cells for controlled molecular study. Here, we report the genomic characterization of a panel of iPSCs from seven 22q11.2 deletion patients and seven matched controls. Multiple iPSC lines were created from each proband, and between one and three lines per proband were analyzed using SNP arrays, for a total of 26 iPSC lines analyzed. We used the new Illumina HumanOmni2.5Exome array which interrogates more than 4.5 million loci genome-wide. The resulting data can be used to detect duplications and deletions with high accuracy. For each iPSC line derived from a patient, we determined the exact extent of the main deletion. We also determined for all lines the overall complement of CNVs and SNPs, including in a small subset of the lines a few additional, smaller CNVs that would have been missed with karyotype analysis, but that are large enough that they should be taken into consideration while using the iPSCs as a model system. The iPSCs are showing full differentiation potential along the neuronal trajectory. For a subset of the iPSCs, we used RNA-Seq analysis and detected gene expression changes in a large number of the genes within the deletion region. These iPSC lines will be highly valuable resource for the analysis of this important microdeletion syndrome, as well as in general for the elucidation of molecular effects of large copy number aberrations on the genomic control of cellular differentiation and functioning. Furthermore our analysis highlights that high-resolution genome analysis in iPSC based model systems should be standard practice.

524W
Common CNVs of Innate Immune Response Genes Defensin DEFA1/A3 And Complement C4A Are Medium Effect-Size Risk Factors for Human Autism Spectrum Disorders (ASD). H. Wang1,2, K. Lintner1, E. Hansen1, B. Zhou1, YL. Wu1, K. Jones1, A. Schwaderer2, D. Hains2, CY. Yu1, G. Herman1. 1) Center for Molecular and Human Genetics the Research Institute at Nationwide Children’s Hospital and Department of Pediatrics, The Ohio State University, Columbus, Ohio 43205; 2) Center for Translation Research the Research Institute at Nationwide Children’s Hospital and Department of Pediatrics, The Ohio State University, Columbus, Ohio 43205.

Products of innate immune response genes are not only crucial in the defense against infections, but are also needed in nerve pruning during the formation of synapses in the central nervous system. Many innate immune response genes are characterized by frequent, inter-individual gene copy number variations (common CNVs). The role of common CNVs for immune-related genes in the pathogenesis of ASD is unknown. Our objective is to determine common CNVs of innate immune response genes for antimicrobial peptides α-defensins DEFA1 and DEFA3, and complement C4A and C4B in genetic risks of ASD. Our study population included 185 ASD patients (90.9% White, 9.7% other races) and >500 race-matched healthy controls from Ohio. Of the White ASD patients, 84% were male and 16% were female; the mean age (±SD) at disease diagnosis was 3.3±1.8 years old. Continuous CNVs for defensins DEFA1/A3 and complement C4 were determined by TaqI genomic restriction fragment length polymorphisms (RFLPs) for all control samples. The relative quantities of DEFA1 and DEFA3 were interrogated by hot-stop PCR and HaeIII-RFLP, and those of C4A and C4B were elucidated by PshAI-Pvull RFLP. The CNV-defined samples facilitated development of sensitive TaqMan-based qPCR amplicons to decipher CNVs for DEFA1/A3 and C4A/C4B in ASD. The copy number for DEFA1/A3 in a diploid genome of White subjects varies from 4 to >15; and for C4, 2 to >6. The mean copy number (±SD) of DEFA1/A3 in ASD is 6.37±1.24, compared to 6.85±1.81 in controls (p=0.0013). On average, ASD patients have 0.5 copy of DEFA1/A3 less than healthy subjects. In particular, there is a marked decrease in the frequency of the high copy-number groups of (≥ copies) in ASD (4.2% versus 13.0%; p=0.0006). The odds ratio (95% CI) for low copy number of DEFA1/A3 in ASD is 3.39 (1.52-7.56). DEFA1 and DEFA3 formation are controlled by the 4B region. The absence of D65 (i.e., homozygous deficiency of DEFA3) has a frequency of 10.8% in ASD, compared to 18.1% in controls [OR=1.81 (1.02-3.21); p=0.035], suggesting the presence of DEFA3 is a risk factor for ASD. For complement C4, we observed a significantly lower copy number of (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copies) in ASD (³ copy...
525T

**TM4SF20** ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities.


1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatric Radiology, Texas Children’s Hospital, Houston, TX; 3) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 4) Northeast Indiana Genetic Counseling Center, Fort Wayne, IN; 5) Division of Developmental Medicine, Vanderbilt University, Nashville, TN; 6) Section of Genetics, Department of Pediatrics, University of California Davis, Sacramento, CA; 7) Tripler Army Medical Center, Honolulu, HI; 8) Center of Excellence in Molecular Genetics of Cancer and Human Diseases, Chulalongkorn University, Bangkok, Thailand; 9) Jan and Dan Duncan Neurological Research Institute at Texas Children’s Hospital, Departments of Pediatrics and Neuroscience, Baylor College of Medicine, TX; 10) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 11) Paw Print Genetics, Genetic Veterinary Sciences, Inc., Spokane, WA; 12) Genome Institute of Singapore, Singapore; 13) The Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam.

White matter hyperintensities (WMH) of the brain are important markers of aging and small vessel disease. WMH are rare in healthy children and, when observed, often occur with comorbid neuroinflammatory or vasculitic processes. Utilizing 180k oligonucleotide-based, exon-focused array-comparative genomic hybridization in 15,493 children, we identified a 4-kb deletion co-segregating with WMH (penetrance ~70%) and early childhood language delay in multiple families, predominantly from Southeast Asia. The 15 unrelated carrier children, mostly of Vietnamese, Thai, Burmese, Filipino, Indonesian, and Chinese ancestry, were characterized for clinical features and WMH. Early language delay, autism spectrum disorders, and/or brain imaging abnormalities. Formal speech and language assessment showed significant discrepancies between verbal and non-verbal skills in the deletion carriers. The premature truncation of TM4SF20 in these families results in a transmembrane protein of unknown function. Minigene analysis showed that the resultant net loss of an exon introduces a premature stop codon, likely leading to the plasma membrane and accumulates in the cytoplasm. This deletion accounts for a strong effect on disease susceptibility related to familial early language delay and autism spectrum disorders in the Southeast Asian pediatric population.

**525W**

Co-evolutionary relationship between chemokines and chemokine receptor: An evolutionary perspective.

C. Tsui1, Y. Hung1, D. Liu2, D. Chen2, 1) La Sierra University, Riverside, CA; 2) Dej et al. was wrong; 3) University of California, Irvine; 3) Center for Stem Cell and Translational Medicine, School of Life Sciences, Aihui University.

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

**526F**

The use of MLPA-based strategy for discrete copy number genotyping of complex multi-allelic CNVs.

P. Kozlowski, M. Marcinkowska-Swajak, ECBIG, Poznan University of Technology, Poznan, Poland.

Copy number variation has recently been recognized as an important type of genetic variation that modifies human phenotypes. Copy number variants (CNVs) are being increasingly associated with various human phenotypes of genetic variation that modifies human phenotypes. Copy number variants (CNVs) are being increasingly associated with various human phenotypes. We used this strategy for discrete copy number genotyping of both simple biallelic and complex multi-allelic CNVs. We used this strategy for discrete copy number genotyping of these extensively studied CNVs, including: CNV-CCL3L1, CNV-DEFB and CNV-UGT2B17, which have been associated with risk of HIV infection, psoriasis and osteoporosis, respectively. Our experiments confirmed the high reproducibility and accuracy of the obtained genotyping results. Acknowledgements: NCN grant 2011/01/B/ZS5/02773.

**527W**

High-resolution analysis of DNA copy number variations in Systemic Lupus Erythematosus patients.

F.B. Barbosa1, M. Simioni2, E.A. Donad2, V.L. Gil-de-Silva-Lopes2, A.L. Simões3, 1) Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil; 2) Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas (UNICAMP), Campinas, SP, Brazil; 3) Division of Clinical Immunology, Department of Medicine, School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil.

Advances in molecular-based techniques for DNA investigation enabled the detection of an important type of genomic variation named copy number variations (CNVs). CNVs are defined as genomic segments, usually greater than 1 kilobase (kb) in size, ranging in copy number when compared to a reference genome. They can contribute to risk variability among individuals in complex diseases etiology. Systemic Lupus Erythematosus (SLE) is an autoimmune disease with strong genetic component characterized by chronic inflammation and autoantibodies production. To date, several loci have been associated with SLE pathogenesis by genome-wide association studies (GWAS). However, there are few analyses about CNVs in SLE patients. The purpose of this study was to determine the role of CNVs in 23 SLE patients. To screen the CNVs, high-resolution array Genomic Hybridization Assay was performed using the Affymetrix® CytoScan™ HD platform. To calculate copy numbers, the data were normalized to baseline reference intensities using 366 samples (270 HapMap samples and 96 healthy normal individuals). Data was analyzed by Chromosome Analysis Suite v.1.2.2 (CHAS) software, which includes the Hidden Markov Model (HMM) algorithm used to determine the copy number states. At total, 406 CNVs were identified (CNV average number per genome was 18), distributed across all chromosomes, except Y. Deletions were more frequent than duplications, 311 and 95, respectively. CNV profile showed 269 CNVs were overlapped by genes, 152 unique CNVs and 59 CNV regions (CNVRs). From all, 39 CNVs detected have not been described in the main database of healthy subjects, the Database of Genomic Variants (DGV). Nine of these have not been described in any structural variants databases. CNVs were found in ten genes previously related with autoimmunity: deletion-type CNVs in STA T4, HLA-DPB2, C4HR, C4H, SNTG1, IL3RA, UGTB15, ADAM3 and duplication-type CNVs in MEC2P and KIAA1267. This is the first report of CNVs in these genes in SLE patients. The identification of CNV in SLE suggests a possible contribution of these variations to development of autoimmunity or the onset of the disease. This is consistent with the observed overlap between CNVs and genes implicated in the development of autoimmunity. The action of these genes in determining SLE or any other autoimmune disease should be investigated in future studies. Support: FAPEA, CAPESES and FAPESP.
Copy number variants (CNV) are one of the major components of human genetic variations and it is thought to contribute to inter-individual differences in diverse phenotypes. Several CNVs have been identified to be associated with systemic lupus erythematosus (SLE) mostly by the target gene approach. However, genome-wide feature of CNVs and their roles in the risk of SLE remain unknown. We aimed to discover SLE-associated CNVs in Korean women. In this study, we performed genome-wide assessments of CNVs and replicated the significant candidate variants in 946 SLE cases and 702 controls. We found that three deletion-type CNVRs in 1q25.1, 8q23.3, and 10q21.3 were significantly associated with SLE. Of the three candidates, CNVRs in 1q25.1 (RABGAP1L) and 10q21.3 were successfully replicated (OR=1.30, P=0.038 and OR=1.90, P=3.6×10−5, respectively) and the associations were confirmed again by deletion-typing PCR. The CNVR in C4 gene, which showed a potential association in the discovery stage, was included in the replication analysis and found to be significantly associated with the risk of SLE (OR=1.88, P=0.01). Through deletion-typing PCR, the exact sizes and breakpoint sequences of the deletions were defined. Individuals with the deletions in all three loci (RABGAP1L, 10q21.3 and C4) had a much higher risk than those without any deletions in all three loci (OR=5.52, P=3.9×10−4) (Kim et al. Arthritis Rheum 2013 65:1055-63). Based on these findings, we develop a multiplex CNV analysis system for predicting the risk of SLE using MLPA-CE-SSCP method. This system contains six CNV targets including the three CNV loci described above. We confirmed the efficacy of our MLPA-CE-SSCP-based lupus risk prediction system.
Homozygous deletions are also mediated by non-LCR structures: AT-rich cruciforms [e.g. t(11;22)] and human endogenous retroviral elements (HERVs) [AZFa deletions in Yq12 and t(4;18)]. Most recently, using chromosomal microarray analysis, we identified nine de novo, identically-sized 3.3 Mb deletions in 3q13.2q13.31, which were likely mediated by HERVs 5,711–6,136 bp in size and sharing stretches of 93-95% DNA sequence identity with their partner. The presence of HERV-mediated deletions, ranging in size between 189 kb and 1.36 Mb and mapping to LCR-free chromosomal regions 1q41, 2p12 (two loci), and 11q24.3 in nine affected subjects, ii) further investigate longitudinally collected samples and describe how the frequencies of variant-clones varies with time within individuals and their relation to onset of disease, iii) to determine which cellular sub-populations are affected by various large-scale aberrations. Our analyses are now extended to include 1153 elderly men from a Swedish population-based cohort UCLA (Uppsala Longitudinal Study of Adult Men) that has been followed clinically for more than 40 years. We aim to i) better describe to frequency and genomic distribution of acquired structural genetic variants in normally aged subjects, ii) investigate longitudinally collected samples and describe how the frequencies of variant-clones varies with time within individuals and their relation to onset of disease, iii) to determine which cellular sub-populations are affected by various large-scale aberrations and to find phenotypic correlations. So far, peripheral blood sampled at ages ranging from 70.7-83.6 years has been analyzed using the illumina 2.5MHumanOmni-chip with strict selection of genotyping quality. The Illumina 2.5MHumanOmni-chip provides for gains and losses, including for gains >10 kb for deletions and CNLNOH. Next generation whole genome sequencing was used for validation. Remarkably, at least one structural change was observed in 42.7% of the genotyped men with gains being the most common type. These results illustrate the high frequency and importance of post-mortem mosaicism in normal cells, which should be studied further for associations with various diseases.

Human Endogenous Retroviral Elements (HERVs) Mediate Multiple Large Deletions and Reciprocal Duplications Suggestive of NAHR. P. Dittwald1, 2, 6, I.M. Campbell2, 6, A. Shuvirkov4, C.R. Beck1, P. Hixson2, T. Gamba1, C.A. Shaw1, A. Gambine1, J.A. Rosenfeld1, P. Stankiewicz1, 2, 5 1) College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warsaw, Poland; 2) Institute of Informatics, University of Warsaw, Warsaw, Poland; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 5) Morsokasski Medical Research Centre, Polish Academy of Sciences, Warsaw, Poland; 6) Equal contribution.

535F
536W

A comprehensive high resolution map of copy number variants shows unique disease risks in a consanguineous Arab population. K. Fakhro1, J.L. Rodriguez-Flores2, N. Younis1, A. Roby1, J.G. Mezey1, 2, R.G. Crystal3 1) Department of Genetic Medicine, Weill Cornell Medical College - Qatar, Doha, Qatar; 2) Department of Genetic Medicine, Weill Cornell Medical College, New York, NY; 3) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

Genetic studies in Arab populations suffer a lack of databases containing population-level background genetic variation, making it difficult to assess the impact of newly discovered mutations in Arab patient cohorts. We present here a genome-wide comprehensive first time analysis of 481 native Qatari, comprising individuals from all 3 Arab subpopulations (Q1 - Bedouin, Q2 - Persian and Q3 - African). We used QuantiSNP and CNVPartition to make CNV calls from illumina 2M SNP array data (mean call rate =99.8%) [1624 high quality CNV Regions (CNVRs) - 1201 deletions, 423 duplications; mean size 49.9 kb; range 59 bp-1.97 Mb; total coverage 21 Mb]. Of 1624 CNVRs, 445 affected 2209 unique genes, including 78 severe Mendelian disease genes. Of interest, 459 of 1624 CNVs (28%) were novel to Qatars; 91 of these affected 243 genes, including 14 severe OMIM genes. In order to also assess CNVs below the detection resolution of arrays, we used CNV calls from whole genome sequencing (WGS) data (>50X coverage) in the same 108 individuals. 400,239 CNVs were detected (average >3,800/individual; size range 50bp-1.2Mb) affecting >246Mb of genomic content - the majority of which were non-genic (>80%) and novel to Qatars (52%). Surprisingly, only a minority (~30%) of array-CNVs overlapped WGS-derived CNVs; thus, despite a >10-fold increase in number of CNVRs called, WGS-algorithms may still miss a significant number of potentially real, large CNVRs in a high proportion of individuals. Cross-referencing these predictions with LCR and their relation to onset of disease, iii) to determine which cellular sub-populations are affected by various large-scale aberrations. Our analyses are now extended to include 1153 elderly men from a Swedish population-based cohort UCLA (Uppsala Longitudinal Study of Adult Men) that has been followed clinically for more than 40 years. We aim to i) better describe to frequency and genomic distribution of acquired structural genetic variants in normally aged subjects, ii) investigate longitudinally collected samples and describe how the frequencies of variant-clones varies with time within individuals and their relation to onset of disease, iii) to determine which cellular sub-populations are affected by various large-scale aberrations and to find phenotypic correlations. So far, peripheral blood sampled at ages ranging from 70.7-83.6 years has been analyzed using the illumina 2.5MHumanOmni-chip with strict selection of genotyping quality. The Illumina 2.5MHumanOmni-chip provides for gains and losses, including for gains >10 kb for deletions and CNLNOH. Next generation whole genome sequencing was used for validation. Remarkably, at least one structural change was observed in 42.7% of the genotyped men with gains being the most common type. These results illustrate the high frequency and importance of post-mortem mosaicism in normal cells, which should be studied further for associations with various diseases.
methods are highly successful at many mCNV loci, including loci with many copy-number CNVs, and their application to GWAS data. We find that these describe new ways to phase and impute the states of multi-allelic, high-resolution segments that segregate in 4, 10, 20 or more different potential copy-number levels in human populations, generally due to tandem or dispersed duplications. Such variants have been refractory to almost all earlier molecular technologies and have therefore not been understood at the levels of alleles, allele frequencies, and haplotypes.

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

To understand how CNV alleles relate to SNPs and haplotypes, we needed to determine not just diploid copy number in each individual, but the contribution of each chromosome and haplotype to diploid copy number. We developed an approach that combines accurate copy number measurements with the information in the diploidy, while maintaining a low false positive rate. Using this method, we were able to confirm and quantitate a previously unreported abundance of somatic genome variation, being present in the form of mosaic lineage-manifested CNVs (LM-CNVs) in human fibroblast tissue. We defined the term LM-CNV to describe CNVs detected by genome-wide analyses in an iPSC line but not in the fibroblast culture from which the given iPSC line was derived -- without making a statement as to the nature of the CNV-forming event. These LM-CNVs had become unmasked in iPSC lines derived in a clonal fashion from the fibroblast tissue of origin. The LM-CNVs had initially been detected by low-coverage whole-genome sequencing in 7 fibroblast samples and 20 corresponding induced pluripotent stem cell lines obtained from two families (Abzov et al., Nature. 2012 Dec 20;492(7429):438–42). We found that on average, an iPSC line has two LM-CNVs. After detecting LM-CNVs by sequencing based analysis in the iPSC lines we investigated the masked, mosaic presence of the same CNVs in the fibroblast tissue of origin. Using standard PCR across the predicted sequence breakpoints we determined that more than half of the LM-CNVs detected in iPSC lines were already present as low allele frequency, mosaic somatic CNVs in the fibroblasts and that up to 40% of fibroblast cells carry such medium-sized to large somatic CNVs. We then used ddPCR and custom designed junction probes to determine the degree of mosaicism of the LM-CNVs. We found that in the samples analyzed the range of mosaicism for LM-CNVs was from 0.3%–14% allelic frequency.
541F Measurement of Cyclin D1 copy number variation at the single cell level using droplet digital PCR. E. Hetzer, Y. Jouvenot, N. Kitgord, K. Hambly, C. F. Novak, R. G. Silliman, M. A. Conti, T. W. Mühleisen, M. M. Nöthen, K.B.Mercer. 1) Department of Psychiatry, Emory School of Medicine, Atlanta, GA; 2) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA.

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

542W Comparison of copy number variation (CNV) calling performance in large numbers of technical replicate SNP array data using three different, widely-used CNV calling algorithms. A. Hofmann1, S. Herrns1, F. Degenhardt1, T.W. Mühleisen1, M.M. Nöthen1, S. Chhatrapati1, S. Hoffmann1,2,3. 1) Institute of Human Genetics Department of Genomics Life & Brain Center University of Bonn; 2) Genomics Research Group Division of Medical Genetics University hospital Basel; 3) Institute of Neuroscience and Medicine, German Cancer Research Center Research Center (DKFZ).

Copy number variants (CNVs) are known to be involved in various cancer types. Accurate copy number variation (CNV) measurement at the single cell level could shed light on the spectrum of CNVs and their clinical relevance. Current methods for single cell CNV include qPCR, sequencing and array CGH. All of these methods require targeted or whole genome enrichment strategies followed by discrete sample measurements. However, the pre-amplification solution creates a different set of biases including increased cost, time and most importantly the potential for skewed results due to bias in the enrichment procedure. Here, we present a method for determining CCND1 copy number state in single cells without the need for pre-amplification. Samples containing subpopulations of cells with various levels of CCND1 amplification were subjected to single cell ddPCR and the results compared to bulk PCR measurements. The results clearly demonstrate the resolving power of ddPCR for the detection of CCND1 copy number state at the single cell level.


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


NextGENe version 2.3.4 includes a sophisticated new algorithm for copy-number variation (CNV) detection from a wide variety of projects, including whole-exome and targeted sequencing panels. This algorithm is based on fitting a beta-binomial model to the coverage ratio. This fitting process results in the amount of noise in the data being measured automatically. The confidence of CNV calls is adjusted based on the amount of noise. Regions are defined for the aligned data in a ‘sample’ project and a ‘control’ project. Whole-exome sequencing can use CDS locations, while targeted sequencing can define regions as the location of amplicons. Each region has a total RPKM coverage (sample plus control) and a coverage ratio (sample divided by total). The fitted equation returns a ‘dispersion’ value for any level of coverage and this value is used to generate beta-binomial distributions for 3 cases: heterozygous deletion, normal (no CNV), and heterozygous insertion (increased copy number). Normalized likelihoods are calculated from these distributions and used in a Hidden Markov Model (HMM) to make the final CNV calls. Each call is given a phred-scored probability score for insertion and deletion.

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

545W Identification of deleterious CNVs in a low SES African American Population. K.B. Mercer1, L.M. Almli2, K.J. Reissler2, J.G. Mullen2. 1) Department of Psychiatry, Emory School of Medicine, Atlanta, GA; 2) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA.

Genome-wide association studies (GWAS) and copy number variant (CNV) discovery have become the standard approaches in attempts to identify genetic variants that result in heritable disorders. The discovery of CNVs has advanced our understanding of genetic variants associated with disorders such as autism and schizophrenia, and will continue to reveal variants that are responsible for yet unknown genetic risk. However, most surveys of CNVs have utilized self-selected case-control populations rather than population-based samples, which may lead to study bias and flawed estimates of CNV prevalence. In the current study, we aim to estimate the prevalence of CNVs associated with genomic disorders in an underserved, impoverished, at-risk African American population. Methods: The Grady Trauma Project is actively recruiting study participants from Grady Memorial Hospital, (Atlanta, GA) which offers healthcare to low income individuals. Study participants are recruited from waiting rooms of either primary care or OB/GYN offices. Willing participants are asked to donate DNA to be used in genetic studies, and to complete questionnaires assessing demographics, trauma exposure and various health outcomes, particularly those related to Depression and PTSD. We used Illumina Omni-Quad genome-wide array data (1M SNPs) derived from 2,927 unrelated African Americans and the PennCNV detection program to identify large CNVs (>100kb) that have previously been found to significantly associate with disease risk. We compared the frequency of CNVs associated with a genomic disorder (n = 64), compared with 1% of controls (n = 88 out of 8,329; p-value = 1.971e-06). The odds ratio (2.22; 95% CI: 1.59-3.10) reveals an increased risk of deleterious CNVs in this population. Notably, we find 2 previously undiagnosed individuals with the 21q22 (VDS2) deletion, and 2 individuals with the 1q21 (CDAD9) deletion. This 2 fold excess implies that this population may be underserved and undertreated with respect to detection of genomic disorders.
546T Determining the utility of MitoExome targeted array CGH in the diagnosis of OXPHOS disorders. H.S. Mountford1,2, E.J. Tucker1; A.G. Compston1,2, N.Lakshmi2, G.E. Verma2,4,3, S.G. Hershman3, V.K. Mootha3,4,5, D.R. Thorburn1,2,6,1) Murdoch Childrens Research Institute, Royal Children’s Hospital, Flemington Road, Parkville, VIC, 3052, Australia; 2) Department of Paediatrics, University of Melbourne, Melbourne, VIC, 3052, Australia; 3) Center for Human Genetic Research and Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Sixth Floor, Boston, MA 02114, USA; 4) Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA; 5) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, MA, 02142, USA; 6) Victorian Clinical Genetics Services, Royal Children’s Hospital, Flemington Road, Parkville, 3052, Australia.

Mitochondrial oxidative phosphorylation (OXPHOS) disorders are the most common inheritable error of metabolism, affecting at least 1 in 5000 live births. OXPHOS diseases are notoriously difficult to diagnose, as they show extreme clinical and genetic heterogeneity, comprising over 150 monogenic disorders with many more ‘disease genes’ yet to be discovered. For about 50% of patients and families, the genetic basis of their OXPHOS disorder remains elusive. Previously, we developed a targeted DNA capture and massively parallel sequencing method to detect variants within the mitochondrial genome and exons of 1034 nuclear genes encoding the mitochondrial proteome or MitoExome1. Our MitoExome study investigated 45 unrelated Australasian patients with an enzyme and clinical diagnosis of an OXPHOS disorder, none of whom had a previous molecular diagnosis. Our studies have now provided a molecular diagnosis in 15 of these patients with mutations in 11 mitochondrial OXPHOS-related disease genes1,2 and identified variants in 2 nuclear disease genes. The remaining 30 patients (60%) had no disease-specific likely pathogenic mutations identified. To further investigate these unsolved patients we designed a Roche NimbleGen 12x135 CGH array to target the MitoExome genes. The design covered all 14,053 exons; 51,380 probes were seen in the European population between CCL4 and CCL3L3. Selection at the CCL3L1 locus in East Asian, West African, European-specific selection at the CCL3L1 locus. The chemokine CCL3L1 binds to the CCR5 receptor, and a 32-bp-deletion mutant (delta32) of CCR5 confers resistance against human immunodeficiency virus (HIV) infection. CCL3L1 copy number (CN) variation has been associated with rheumatoid arthritis and susceptibility to HIV infection, supporting the hypothesis that genetic variants in CCL3L1-CCR5 have previously been subject to selection. We therefore examined selection at the CCL3L1 locus in East Asian, West African, European and American populations using 3 megabases of 1000 Genomes Project data in this region. Two tests based on extended haplotype homozygosity were applied: Integrated Haplotype Score (iHS) and Cross Population Integrated Haplotype Score (cHS). The iHS test identified a consistent window for applying therapies to newborn babies is very small as >95% of motor neurons are lost within 6 months. The promise of reducing the severity of the disease enhances the importance of early diagnosis and intervention before irreversible motor neuron damage occurs. The autosomal recessive nature of the disease provides the possibility that through careful screening the incidence of the disease could be significantly reduced. Traditional carriers (1:0) can be identified using qPCR, however this methodology fails to identify carriers that have cis-configured copies of SMN1 (2:0). Here, we use droplet digital PCR (ddPCR) to accurately quantify the number of SMN1 and SMN2 genes in 384 HapMap samples from four different ethnic backgrounds. Furthermore, we employ digital linkage analysis to determine the frequency of SMN1-to-SMN2 conversion and vice versa. This information is useful for understanding the factors that influence the frequency of carriers, which is estimated to be 1/50. Lastly, we explore the possibility and limitations of using linkage analysis for identifying cis-configured carriers, which are missed by all other methodologies.

547F European-specific selection at the CCL3L1 locus. H.T. NGUYEN1,2, T.R. MERRIMAN1, M.A. BLACK2, 1) Biochemistry Department, University of Otago, 710 Cumberland street, Dunedin 9016, New Zealand; 2) Mathematics and Statistics Department, University of Otago, 710 Cumberland street, Dunedin 9016, New Zealand.

Infectious disease has been a prominent selective agent in Europe for over one thousand years. The chemokine CCL3L1 binds to the CCR5 receptor, and a 32-bp-deletion mutant (delta32) of CCR5 confers resistance against human immunodeficiency virus (HIV) infection. CCL3L1 copy number (CN) variation has been associated with rheumatoid arthritis and susceptibility to HIV infection, supporting the hypothesis that genetic variants in CCL3L1-CCR5 have previously been subject to selection. We therefore examined selection at the CCL3L1 locus in East Asian, West African, European and American populations using 3 megabases of 1000 Genomes Project data in this region. Two tests based on extended haplotype homozygosity were applied: Integrated Haplotype Score (iHS) and Cross Population Extended Haplotype Homozygosity (cHS). The cHS test identified a consistent window for applying therapies to newborn babies is very small as >95% of motor neurons are lost within 6 months. The promise of reducing the severity of the disease enhances the importance of early diagnosis and intervention before irreversible motor neuron damage occurs. The autosomal recessive nature of the disease provides the possibility that through careful screening the incidence of the disease could be significantly reduced. Traditional carriers (1:0) can be identified using qPCR, however this methodology fails to identify carriers that have cis-configured copies of SMN1 (2:0). Here, we use droplet digital PCR (ddPCR) to accurately quantify the number of SMN1 and SMN2 genes in 384 HapMap samples from four different ethnic backgrounds. Furthermore, we employ digital linkage analysis to determine the frequency of SMN1-to-SMN2 conversion and vice versa. This information is useful for understanding the factors that influence the frequency of carriers, which is estimated to be 1/50. Lastly, we explore the possibility and limitations of using linkage analysis for identifying cis-configured carriers, which are missed by all other methodologies.

548W Integration of copy number and structural variation across species can provide unique insight into disease pathology: Osteogenesis Imperfecta and Autism cases explored. A. O'Hara, L. Culot, S. Verma, Z. Che, S. Shams. BioDiscovery Inc., Hawthorne, CA.

Rapid identification of candidate genomic aberrations responsible for disease phenotype is important in both research and clinical settings. Increasingly, structural variant data must also be combined with sequencing data in order to uncover pathogenic events and then be further resolved through comparison with related genomes. Here we explore two constitutional cases, Osteogenesis Imperfecta in a proband with consanguineous parents, explained through comparison with sibling, and an autism case with a suspected de novo SHANK2 deletion and family trio data. An approach that can integrate structural and sequence variation across related samples will be presented.
550F

Frequency of gene usage and copy number variation within the rearranged Immunoglobulin Heavy-Chain Variable locus based on immune repertoire sequencing. M.J. Rieder1, D. Williams1, A. Sherwood1, R. Emerson1, C. Desmarais1, M. Chung1, H. Robins1,2, C. Carlton1,2. 1) Adaptive Biotechnologies, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA.

The human adaptive immune system is composed of both B and T cells that undergo somatic recombination at specific loci to create rearrangements of Variable (V), Diversity (D) and Joining (J) gene segments. For the B-cell immunoglobin receptor heavy-chain (IGH), the CDRI3 regions are defined by the VDJ gene segments and nucleus insertions/deletions at these junctions that create the vast sequence diversity of the IGH repertoire. Characterizing the germline DNA in these regions is impeded by the high sequence similarity between gene segments, mutation and copy-number variation (i.e. large insertions/deletions). Currently, there is a fundamental lack of information about the baseline IGH immune repertoire V gene usage and diversity within healthy human controls. To provide an estimate of this, we sequenced functionally recombined gene segments to infer the underlying gene structure. From a set of 132 healthy controls we sorted C19+/CD27+ B-cells from whole blood and amplified genomic DNA using a highly multiplexed PCR assay that targeted the rearranged IGH receptor locus. Following DNA sequencing and data processing to assign V, D and J gene families and names, we examined the usage frequency of IGHV gene segments across all individuals. We found that of the 98 V gene segments only 56 (57%) were used at a frequency > 0.1%, and ~10 showed little to no usage (present in <1% of individuals). This data also allowed us to identify two IGHV genes currently annotated as orphans (pseudogenes assigned to an alternate chromosomal location) that had previously functional usage (IGHV4/OR15-8; IGHV3/OR16-09) and therefore must reside at the IGH locus on chromosome 14. Finally, by taking this functional approach we were able to screen all V gene segments for germline copy-number variation (e.g. large insertion/deletion events encompassing individual genes) by looking for an excess of deletion events or modal changes in gene usage. We confirmed that existence of 12 of 15 previously identified deleted IGHV gene segments. Strong deletion evidence was observed for an additional 37 V gene segments (IGHV3-33, IGHV4-04, IGHV4-41, IGHV3-35) and ten with highly likely germline deletion events. These data suggest that functional immune profiling of rearranged immune receptors provides a more robust method of identifying individual structural variation and provides insight into the immune repertoire of healthy controls.

550T

Comprehensive comparison of copy number variations detection using Illumina Omni 2.5M and Affymetrix CytoScan® arrays. C. TAM1, E. WONG1, H. GUIT1, S. CHERRY1,2, P. SHAM1,2,4, P. TAM1,2,5, M. GARCIA-BARCELO1,2,3. 1) Department of Psychiatry; 2) Centre for Genomic Sciences; 3) Department of Surgery; 4) State Key Laboratory of Brain and Cognitive Sciences; 5) Centre for Reproduction, Development and Growth, the University of Hong Kong, Pokfulam, Hong Kong.

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

553F

Characterisation of the RNU2 CNV, a bulky neighbour for BRCA1. C. Teixeira1,2, N. Monnet1, M. Imbert1, M. Buisson1, L. Barjhoux1, C. Cuenin2, C. Schluth-Bolarid3,4, D. Sanjaville4,4, Z. Herceg5, E. Conseiller5, M. Ceppi6, L. Duret5, OM. Sinilnikova1,2, S. Mazoyer1. 1) Genetics of Breast Cancer, Cancer Research Center of Lyon, CNRS UMR5286/Inserm U1052/Université Lyon 1, Lyon, France; 2) Genomic Vision, Bagneux, France; 3) Service de Génétique, Laboratoire de Cytogénétique Constitutionnelle, Centre de Biologie et de Pathologie Est, Hospices Civils de Lyon, France; 4) INSEM U1028, CNRS UMR5292, Université Claude Bernard Lyon 1, Equipe TIGER, 69000 Lyon, France; 5) Laboratoire de Biométrie et Biologie Évolutive, Université de Lyon, Université Lyon 1, CNRS, INRIA, UMR5558, Villeurbanne, France; 6) Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon, Centre Léon Bérard, Lyon, France; 7) Epigenetics Team, Molecular Carcinogenesis and Biomarkers Group, International Agency for Research on Cancer, F-69008, Lyon, France.

The question of the implication of multiallelic CNVs in complex traits remains largely open as most of them cannot be genotyped by array technology. In this work, we focused on the RNU2 locus, a variable number of tandem repeats that contains the gene coding for the snRNA U2, an essential element of the splicing machinery. RNU2 was shown many years ago to reside close to the breast cancer susceptibility gene BRCA1 but is still missing from the latest human genome assembly and cannot therefore be investigated by recent genomic approaches. Using unassembled contigs, we precisely located RNU2 within the chromosome 17 reference assembly, 124 kb telomeric of the breast cancer susceptibility gene BRCA1. By FISH analyses on combed DNA (Molecular Combing), we determined more precisely the exact allelic number of repeats than with the previously used Pulse Field Gel Electrophoresis technique and found a range of 6-82 and a level of heterozygosity of 98% in 41 individuals. We used the 1,000 Genome Project data for analysing the variability of this macrosatellite by mapping sequence with unlocalized human genomic contigs and confirmed its high degree of polymorphism, suggesting that depth-of-coverage calculation is a very useful tool for accurate CNV characterization. We found 24 frequent SNPs within the RNU2 basic unit, and the genotype data for 1,106 individuals confirmed previous results showing a concerted evolution of this CNV. Thanks to our precise location, we were able to confirm that the RNU2 array is within the BRCA1 linkage disequilibrium block, which allowed us to study the RNU2 array transmission over a large number of generations. Our resulting observation is that a highly variable locus can nevertheless be highly stable. Given the high level of polymorphism of this locus, we also measured the expression of U2 snRNA by qRT-PCR in 16 individuals carrying 36 to 110 copies of RNU2 and found no correlation with the number of copies. Interestingly, we also measured methylation level of RNU2 and found a higher level of methylation of the RNU2 CNV in individuals with the highest copy number, suggesting that methylation could be involved in dosage compensation. These findings extend our knowledge of a recently neglected CNV that could be valuable for evaluating the potential role of structural variations in disease due to its location next to a major cancer susceptibility gene.
554W
Testing rare coding deletions identified using dense exome chip array data for contribution to type 2 diabetes. M. Thummer1, A. Mahajan1, N. Robertson1, A. Kumar1, W. Rayner1, F. Karpe1, C. Palmer1, T. Spector1, M. McCarthy1, K. Gaulton1. GoT2D consortium. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, UK; 3) Diabetes Centre, Biomedical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Ninewells Hospital, Dundee, UK; 4) Department of Twin Research, King’s College London, London, UK.

Loss of function (LoF) of gene activity is a major contributor to phenotypic variability and to both Mendelian and complex disease. To what extent to which LoF events might influence susceptibility to type 2 diabetes (T2D), however, is currently unknown. In particular, partial or complete deletions of coding sequence leading to gene LoF have not been explored on a genome-wide scale.

We investigated exonic deletions by using the Illumina HumanExome SNP array (consisting of >54,000 primarily exonic variants). We developed an analytical pipeline for deletion discovery from this array that applies several existing structural variant calling algorithms (PennCNV and QuantiSNP) and merges the resulting calls together to produce one set of calls per sample. We then applied the same procedure to exome chip data from 11,686 UK T2D cases and 6,323 controls, with a total of 372 deletions and 62 duplications identified. Of which were low frequency (MAF <0.01) and 48% (100) were singletons. We tested these 208 variants for individual association to T2D at each single SV, and identified large deletions on chromosome 7 overlapping RARG, 7p11.2 and CNE-2, and on Xp11.23 and CNE-3 which means that the shortest region of overlap for these three deletions only includes CNE-3. Even more remarkable is that we also found a duplication encompassing CNE-5, CNE-3 and CNE-2 in 1 patient with ISS. These observations therefore point to a possible crucial role for CNE-3 as enhancer element in the regulation of SHOX.

554F
Study of 455 molecularly unsolved LWD and ISS cases: identification of two deletions and the first duplication upstream of SHOX. H. Verdin1, L. Borns1, E. Debaere1, B. Dehaene1, G. Matthijs1, E. Maria2, S. DePoorter2, E. De Baere1. 1) Center for Medical Genetics, Ghent University, Ghent, Belgium; 2) Center for Human Genetics, University of Leuven, Leuven, Belgium; 3) Department of Pediatrics, Ghent University Hospital, Ghent, Belgium; 4) Department of Pediatrics, University of Saint-Jan, B. Edegem, Belgium.

Short stature homeobox-containing gene (SHOX) is located in the telomeric pseudo-autosomal region (PAR1) on the short arms of the X- and Y-chromosome. This PAR1 region also contains several enhancer elements, five of which are found within the first 100,000 bp of chromosome 15 overlapping TSPAN3 and PSTPIP1 (37,500bp, P=1.1×10−4) with significant T2D association after correction for multiple testing. Both are present in the Database of Genomic Variants (DGV), supportive of these being true deletion events. We further examined whether these structural variations are significantly associated with copy number of the gene. We suggest that the CNV may affect signaling upon LILRB3/A6 ligation. Furthermore, this study adds strength to the recognition that the PAR1 region contains several enhancer elements.

555T
Multiplex emulsion haplotype fusion PCR to determine haplotypes at structurally complex regions. J. Tyson, H.A. Black, J.A.L. Armour. School of Life Sciences, University of Nottingham, Nottingham, United Kingdom.

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

555W
Diversity of the human LLR3B/AL36 locus encoding a myeloid inhibitory and activating receptor pair. N. Vinc1,2, A. Bashirova3, E. Maris2,3, E. Mochalova2, X. Yu1, M. Carlington1,2, 1) Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIG-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD; 2) Ragon Institute of MGH, MIT and Harvard, Boston, MA; 3) University of Maryland, Baltimore County, Baltimore, MD.

Leukocyte Ig-like receptor B3 (LLR3B) and LLR3A6 represent a pair of inhibitory/activating receptors with identical extracellular domains and unknown ligands. LLR3B can mediate inhibitory signaling via ITIMs in its cytoplasmic tail whereas LLR3A6 can signal through association with an activating adaptor molecule, FcRγ, which bears a cytoplasmic tail with an ITAM. The receptors are encoded by two highly polymorphic neighboring genes within the Leukocyte Receptor Complex (LRC) on human chromosome 19. Here we report that the two genes display similar levels of single nucleotide polymorphisms with the majority of polymorphic sites being identical. In addition, the LLR3A6 gene exhibits copy number variation (CNV) and does not affect gene expression. A screen of healthy Caucasians indicated that 32% of the subjects possessed more than 2 copies of LLR3A6, whereas 4% have only one copy of the gene per diploid genome. Analysis of mRNA expression in the major fractions of PBMCs showed that LLR3A6 is primarily expressed in monocytes, similarly to LLR3B, and its expression level correlates with copy number of the gene. We suggest that the LLR3A6 CNV may influence the level of the activating receptor on the cell surface, potentially affecting signaling upon LLR3B/A6 ligation.
Highly variable tandem repeat genes: hotspots for primate evolution

559T

559F

Copy number variations (CNVs) polymorphisms are common in phenotypically normal individuals and some have been shown to increase risk for human diseases, for example schizophrenia and other neuropsychiatric disorders. The ability to detect single nucleotide polymorphisms (SNPs) as well as copy number variation in the same genome screen is efficient and advantageous. We show that Affymetrix® Axiom® arrays, designed to detect genome-wide associations with SNPs and InDels, also detect copy number variations. The CNV detection method computes log2(ratio) for each marker, where for each marker site the sum of allele intensities from an individual is normalized by the intensity produced by taking the average of the two alleles. Together, these findings demonstrate that previous GVAS have had limited power to assess the impact of TR gene variation in disease susceptibility. To address these shortcomings, we are currently screening 26 of the most polymorphic TR genes in populations with various autoimmune diseases for which a majority of disease variance remains uncharacterized.

560W
Fusion genes resulting from complex duplications in chromosome Xq28. L.W. Zuccherato4, B. Alleva3, C.M.B. Carvalho3, J.R. Lusksi1,2,3. 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX.

Changes in gene dosage likely are responsible for many genomic disorders. However, advances in technology are revealing that complex genomic rearrangements (CGRs) can lead to the formation of new genes that have potential effects in the disease phenotype. Chromosomal rearrangements can be produced by recombination and replication mechanisms and can convey diverse phenotypic effects. Gene fusions have been observed in somatic alterations such as cancer, and despite the importance of these fusion genes, the underlying mechanism(s) that result in their formation in constitutional chromosome alterations and other rearrangements are poorly understood. Data from the breakpoint junction sequencing of two individuals carrying duplications involving the Xq28 region enabled us to predict a complex pattern of rearrangements that include the fusion of exons from genes F8/CSAG1 and BCA3P1/TEX28. RT-PCR experiments confirmed the expression of the new fusion genes in transformed lymphoblastoid cell lines from these patients. The rearrangement found in the F8/CSAG1 produced an inversion of the 5'UTR of the F8 gene, leading to a transcript in the same orientation as transcript CSAG1. Moreover, the newly expressed F8/CSAG1 transcripts included a partial sequence of ERVL-MaLR intrinsic repetitive element. This provides evidence that the 'exonization' of repetitive elements can occur in the formation of a fusion gene and may be triggered by changing the genomic context in which a gene is transcribed. Future experiments will address the functional relevance of these changes at the protein level and the further impacts of structural variation on patient phenotypes. Thus, complex rearrangements mediated by the replication mechanism FoSTeS/MMIR may contribute to exon shuffling processes and diversify the repertoire of expressed transcripts, and therefore CGRs may have a role in the evolution of both individual genes as well as the human genome.

561T

Genetic variation results in human population diversity and differential disease susceptibility. Different methods, such as quantitative PCR, array comparative genomic hybridization (aCGH), Sanger or next-generation sequencing have been used for identification of these polymorphic loci, however these methods may have either low sensitivity in repetitive regions, are too labor-intensive and time consuming, or do not provide positional information about complex rearrangements. To address these limitations, we use a single-molecule mapping approach to identify structural variation as well as to provide scaffolds for de novo assembly of a diploid human genome, NA12878, from 'short reads' sequences. Genome mapping utilizes highly parallel nanochannel arrays in which thousands of very long single DNA molecules are linearized and imaged. This novel approach is automated on the Irys System, which can scan the entire genome rapidly to generate physical maps that provide a more comprehensive view of the genome. NA12878, the daughter in a CEPH-CEU trio, was used as this sample has been genotyped and sequenced extensively as part of the International HapMap and 1000 Genomes Projects. We nicked and fluorescently labeled DNA fragments, size ranging from 100 kb - 500 kb, at Nt. BspQI (GCTTTCN) sites. To date, we generated over 50X coverage data and constructed de novo assembled genome maps covering more than 90% of the genome using an automated assembly pipeline. We identified many structural variants (indels) including those previously known in this sample. Overall, this genome mapping approach is simple and can be performed in any modern molecular genetics laboratory.
The release of Phase I of the 1000 Genomes Project (GWAS) has been a major source of variation in the human genome, but can be challenging to assay directly with current chip-based genotyping methods used for genome-wide association studies (GWAS). Therefore, there is a need to improve our ability to detect small CNVs in our database were previously reported to undergo Alu-mediated CNVs. We cross-referenced this list of loci with potential Alu-mediated CNVs identified in over 25,000 patients from our diagnostic laboratory. We discovered a significant enrichment in small (<250 kb) genomic deletions involving the genes with the largest numbers of these distinct Alu-Alu pairs. Notably, a number of these genes are known to cause genome instability and may allow prediction of disease causing CNVs at a genomic level. Small CNVs in our database were previously reported to undergo Alu-mediated CNVs. We cross-referenced this list of loci with potential Alu-mediated CNVs identified in over 25,000 patients from our diagnostic laboratory. We discovered a significant enrichment in small (<250 kb) genomic deletions involving the genes with the largest numbers of these distinct Alu-Alu pairs. Notably, a number of these genes are known to cause genome instability and may allow prediction of disease causing CNVs at a genomic level. Small CNVs in our database were previously reported to undergo Alu-mediated CNVs. We cross-referenced this list of loci with potential Alu-mediated CNVs identified in over 25,000 patients from our diagnostic laboratory. We discovered a significant enrichment in small (<250 kb) genomic deletions involving the genes with the largest numbers of these distinct Alu-Alu pairs. Notably, a number of these genes are known to cause genome instability and may allow prediction of disease causing CNVs at a genomic level.
The advent of whole genome sequencing has allowed for population genetic analyses, previously only theory, to be applied to the data they were designed for. In particular, genome sequences from multiple populations across the globe allow for inferences to be made about historical population events. Utilizing a method originally proposed by RC Griffiths (1), and later leveraged by Johnston and Cutler (2), the effective population size of the population an individual is drawn from can be calculated. In this approach, genetic analyses, previously only theory, can now be applied to the data they were designed for. This deficiency, we have developed an approach to discover and genotype previously undiscovered numt insertions using whole genome, paired-end sequencing data. We have identified more than 150 novel sites of numt insertions by applying our method to over a thousand individuals in twenty populations from the 1000 Genomes Project and other datasets. This expands our current knowledge of existing numt locations in the human genome by 20% and represents a substantial increase from the 14 polymorphic numt loci previously reported. Over 90% of the newly identified numts were found in less than 1% of the samples we examined to date, suggesting that they occur infrequently in nature or have been rapidly removed by purifying selection. We further extended our analysis to the whole genomes of over sixty non-human primates belonging to four genera and collectively examined these variants for biases in sequence context at insertion sites and compared with human genomic sequences. We believe this research will help clarify to what extent that numts play a role in hominoid diversification and phenotypic variability.

Objective: Genome wide association studies (GWAS) have identified various migraine susceptibility variants. We aim to replicate five GWAS associated polymorphisms (rs1835740, LR1P rs11172113, TRPM6 rs10166942, PRDM16 rs2651899 and TGFBR2 rs7640453) in North Indian population. Furthermore, we checked the SNPs in strong linkage disequilibrium (LD) with the selected variants. We also undertook to predict the functional effect (in silico) of the variants. Design: The study included 340 migraineurs and 200 controls. Genotyping was performed by PCR-RFLP, ARMS PCR and Taqman. Logistic regression was used for association analysis. LD plot was prepared using genotyping data retrieved from ENCODE and HapMart. Functional effect was predicted by F-SNP and FastSNP. Results: We did not observe any significant effect of the variant genotype or allele of the first migraine GWAS associated marker, rs1835740. However, significance was observed in case of heterozygous genotype for total migraineurs and in few subgroups: migraine without aura (MO) and females. We suggest potential protective effect of LRPI rs11172113 and some of the SNPs in strong LD were predicted to affect transcriptional regulation. Functional effect of LRPI rs11172113 could not be found but SNP in strong LD with it was found to affect transcription factor binding sites. Conclusion: We present the first replication study of GWAS associated polymorphisms in a population other than European.
569W

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

571F
Common variation in the Melanocortin 4 Receptor gene (MC4R) is associated with increased food intake and obesity in American Indians. Y.L. Muller, R.L. Hanson, M. Thearle, D. Hoffman, B. Gona, K. Huang, S. Kobes, S. Votruba, J. Krakoff, W.C. Knowler, C. Bogardus, L.J. Baier, Diabetes Molecular Genetics Section, National Institutes of Health, Phoenix, AZ.

We previously identified 10 rare non-synonymous mutations in the Melanocortin-4-Receptor (MC4R) gene by sequencing 900 whole genomic DNA samples lives in the Gila River Indian Community of Arizona, where most of the residents are of Pima Indian heritage. The overall prevalence of MC4R functional coding nucleotide polymorphisms (SNPs; N=63) spanning a ~414 kb region encompassing MC4R were genotyped in 5880 American Indians who had longitudinal measures of body mass index (BMI). Most of these individuals had a measure of maximum BMI from a non-diabetic exam during childhood (maximum z-score, age and sex adjusted, between ages of 5 and 20 years) and had a measure of maximum BMI from a non-diabetic exam during adulthood (maximum BMI after the age of 15 years). A subset of these individuals had been studied as inpatients in our Clinical Research Center and had measures of body composition as assessed by dual x-ray absorptiometry (N=415) and ad libitum food intake measured over 3 days by an automated vending machine system (N= 203). A promoter SNP (rs6097783) with a risk allele frequency (RAF) of 0.48 was associated with BMI in childhood (p=0.01 for age and sex adjusted Z score) and adulthood (p=0.0007, adjusted for age, sex, birth year and ethnicity). Another promoter SNP (rs11872992; RAF=0.94) was also associated with BMI in childhood (adjusted p=0.005) and adulthood (adjusted p=0.06), where the risk allele was further associated with a higher percentage of body fat (p=0.01, adjusted for age, sex and family membership) and an increase of 640 kcal/day in total food intake (p=0.003, adjusted for age, sex, percent body fat, family membership and ethnicity). The rs11872992 promoter variant was predicted to cause a loss of function by the Ingenuity Variant Analysis. In vitro luciferase assay of rs11872992 found that the risk allele had a modest decrease in promoter activity (p=0.005). We conclude that both common and rare variation in the MC4R gene contribute to risk of childhood and adulthood obesity in American Indians.

570T

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

572W
Next-generation sequencing of complete mitochondrial genomes of Slovenian Leber’s Hereditary Optic Neuropathy patients revealed one novel mutation and several probable synergistic variants. D. Glavac1, M. Tajnik1, M. Jarc-Vidmar2, M. Hawlina2. 1) Department of Molecular Genetics, Faculty of Medicine, University of Ljubljana, Korytkova 2, SI-1000 Ljubljana; 2) Eye Hospital, University Medical Centre, Grabljevoˇeva 46, SI-1000 Ljubljana.

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

Posters: Genome Structure, Variation and Function
573T Genetic diversity in black South Africans from Soweto. M. Ramsey, S. Hazehurst, Y. Li, M. Waldvogel, J. Eichenberger, S.A. Norris, M. Govind, A. May. 1) Division of Human Genetics, University of the Witwatersrand, Faculty of Health Sciences, Johannesburg, South Africa; 2) National Health Laboratory Service, Johannesburg, South Africa; 3) Wits Bioinformatics, University of the Witwatersrand, Johannesburg, South Africa; 4) Novartis Institutes for Biomedical Research (NIBR), Human Genetics and Genomics, Basel, Switzerland or Cambridge, MA, USA; 5) MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, University of Health Sciences, Witwatersrand, Johannesburg, South Africa; 6) Division of Rheumatology, Chris Hani Baragwanath Hospital and the School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

Background: Due to the unparalleled genetic diversity of its peoples, Africa is attracting growing research attention. Several African populations have been assessed in global initiatives such as the International HapMap and 1000 Genomes Projects. Notably excluded, however, is the southern Africa region, which is inhabited predominantly by south eastern-Bantu speakers, currently suffering under the dual burden of infectious and non-communicable diseases. Limited reference data for these individuals hampers medical research and prevents thorough understanding of the underlying population substructure. Here, we present the most detailed exploration, to date, of genetic diversity in 14 unrelated south eastern Bantu-speaking South African residents in urban Soweto (Johannesburg). Results: Participants were typed for ~4.3 million SNPs using the Illumina Omni5 beadchips. PCA and ADMIXTURE plots were used to compare the observed variation with that seen in standard populations worldwide. While the Sowetans and other south eastern Bantu-speakers, are a clearly distinct group from other African populations previously investigated, reflecting a unique genetic history with small, but significant contributions from diverse sources. To assess the degree of sample relatedness, we compared our results to a larger control and case sample selected for a rheumatoid arthritis study. The control group showed good clustering with our sample, but among the cases, individuals who demonstrated notable admixture of possible clinical relevance. Conclusions: Our data represent a suitable reference set for south eastern Bantu-speakers, on par with a HapMap type reference population, and constitute a prelude to the Southern African Human Genome Programme.

573W The completion of the 1000 Genomes Project. A. Auton, The 1000 Genomes Project Consortium. Albert Einstein College of Medicine, Bronx, NY.

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

Through these efforts, nearly all the short sequence variants detected in a typical human genome sequenced today will have been previously identified. To demonstrate the utility of the 1000 Genomes resource, we have obtained deep whole genome sequence data from 136 diverse trio's, allowing direct comparison of the high and low depth sequencing approaches, as well as investigation of integrated calling methods. We show that even for high coverage genomes, utilizing 1000 Genomes data can improve the accuracy of detected variation and aid downstream interpretation.

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


Background: The African Genome Variation Project (AGVP) is an international collaboration that aims to produce a comprehensive catalogue of human genetic variation in Sub-Saharan Africa (SSA) in order to extend our understanding of population history, genetic diversity among populations in SSA, as well as provide a global resource to help design, implement and interpret whole genome sequencing in SSA. Phase I of the AGVP generated a catalogue of 100 unrelated individuals from each of 16 ethno-linguistic groups from SSA with the 2.5 million (M) Illumina genotype arrays. Here, we present preliminary data from phase II of the AGVP. This project specifically aims to extend and integrate data from Phase I by whole genome sequencing (WGS) of 2000 individuals from genetically diverse populations within SSA in order to develop a resource that provides a comprehensive catalogue of genetic variation in SSA along with core phenotypes accessible to the global scientific community. Methods: AGVP phase II has generated deep (~60X) whole genome sequence data from genomes of 3 geographically distinct populations from Ethiopia, South Africa and Uganda. By October 2013, we aim to extend sequencing to include 1000 individuals from distinct populations including Khoe-San groups in Namibia. WGS was carried out using Illumina Hiseq 2000. Read alignment, realignment and genotype calling was carried out using GATK best practice recommendations. Results: We identified 19M, 19.2M and 18.8M variants across the whole genome among Baganda (100 samples), Ethiopia (120 samples) and Zulu (100 samples) population groups. Of these, 10%, 18.1% and 9% respectively were novel, with a higher percentage of non-synonymous and non-synonymous variants enriched in the lower allele frequency spectrum, with the ratio of non-synonymous to synonymous SNPs being reversed, indicating enrichment for functional variants. The overlap of variants between populations was low, with only ~5% of variants being private in these populations, respectively. Genetic differentiation measured by Fst was high among populations (mean pairwise Fst=0.536), with differentiation being greatest between Ethiopian and other populations. Conclusions: We identified a large number of novel and potentially functional rare variants among the 3 populations sequenced, with a high proportion of private variants in each population. These findings argue for more detailed characterisation of genetic variation in SSA populations using WGS approaches.

574T Quality control metrics for whole-exome sequence data analysis and detailed genetic variation from 6,000 disease variants within the UK10K project. J. Floyd on behalf of the UK10K Consortium: exomes. Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The UK10K project is a large-scale next generation sequencing project that has sequenced nearly 4,000 low-coverage whole genomes (~6X depth) and over 5,000 high-coverage whole exomes (~50X depth). Very few projects of this magnitude have been undertaken to date. Therefore, detailed characterisation of the developed quality control (QC) and variation observed with comparison to other large resources such as the 1000 Genomes Project and the NHLBI EVS is of benefit to the broader community. Genome-wide association studies have previously demonstrated the importance of high quality data prior to analysis and these issues are even more likely to influence analysis of next-generation sequence data. The sample-level QC undertaken included identifying sample contamination (~3% of samples sequenced), cryptic population relatedness (0–3% depending on subgroup), genuine familial (but sometimes inaccurately reported) relatedness, and their respective abilities to accurately call singleton variants. To detect possible effects of changing sequencing chemistry and/or variant calling we also report metrics such as the Ti/Tv ratio for SNVs and 3n/non-3n ratio for indels, mean depth and mean number of variants by minor allele frequency bins. After sample QC, 5233 samples remained for analysis. Ethnic outliers and related samples were removed or controlled for within the analyses. For instance, non-UK families were intentionally gathered and analyzed for some of the rare diseases. On average, nearly 70,000 variants passed variant QC for each of the 5 datasets. Variant QC included VQSR and a per-subject variant filter based on genotype quality. After sample and variant QC, we report a list of highly mutable genes where we see a higher density of variants with a particular expansion in sampling from populations with African or South Asian ancestry. In total, low coverage (~5X) whole genome sequence data have been combined with high coverage (~80X) exome data for over 2,500 individuals from 26 global populations.

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

576T Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

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

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

Through these efforts, nearly all the short sequence variants detected in a typical human genome sequenced today will have been previously identified. To demonstrate the utility of the 1000 Genomes resource, we have obtained deep whole genome sequence data from 136 diverse trio's, allowing direct comparison of the high and low depth sequencing approaches, as well as investigation of integrated calling methods. We show that even for high coverage genomes, utilizing 1000 Genomes data can improve the accuracy of detected variation and aid downstream interpretation.

In addition, the project has also been working towards ascertaining additional types of variation. While earlier phases of the project focused on simple types of genetic variation in easily accessible parts of the genome, the final phase is using novel analytic methods to identify more complex types of variation such as multi-allelic polymorphisms, complex structural variants, and short tandem repeats (STRs). By integrating multiple detection approaches that leverage information from read mapping, locally reassembled, and full-scale de novo assembly, we are able to generate a more complete picture of human genetic variation.
577F
Detailed annotated whole genome sequences of a Tibetan Trio family revealed high-altitude genetics and demographic insights. M. He1, A. Asan1, X. Jin1, E. Huerta-Sanchez2, R. Wang2, Z. Cuo2, Y. Shan1, Y. Shi1, M. Yang1, X. Xie1, K. Harris3, D. Cao1, I. Song1, J. Zhao1, Z. Su1, J. Zhang1, Y. Chang1, C. Yu1, H. Huasang1, J. Luosang1, X. Yi1, Y. Liang1, R. Nielsen2, Jun. Wang1, W. Wang1, Jian. Wang1. 1) BGI-Shenzhen, Shenzhen, Guangdong, China; 2) Statistics and Integrative Biology, UC Berkeley, Berkeley, CA; 3) Naqu local Peoples Hospital, Lhasa, China; 4) The Second People’s Hospital of the Tibet Autonomous Region, Naqu, China.

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

578W

Deep whole genome sequencing across multiple samples provides an opportunity to elucidate the genetic structure and diversity of a population at finer resolution. We conducted deep sequencing of 38 cosmopolitan Indians in Singapore to understand the genetic structure of Asian Indians which is unique in relation to its historical background and genetic admixture. Genetic diversity of Asian Indians is a powerful avenue to establish disease loci map. Furthermore, this population which contributes close to one fifth of the world’s population is absent in the Phase 1 release of the 1000 genomes project. Therefore, our aim of sequencing 38 Asian Indians is to analyze their genetic structure and fully characterize their genetic profiles.

On top of that, the formal study of their genetic diversity would reveal many implications in medical research and serve as an invaluable resource. Given the availability of whole genome sequences of archaic hominin genomes (Denisovan and Neandertal) and 15 other populations across the globe, we examined the relationship of modern human genomes with archaic hominin genomes. The genetic makeup of our samples was more closely related to Southern Indians than Northern Indians. We then demonstrated that 38 Indian genomes has high genetic sharing with Denisovan, evidence of potential admixture between Denisovan and modern humans who are ancestors to Asian Indians. Moreover, we discovered genetic ancestry of mitochondrion and Y chromosome that deviates from the autosomes.

579T
Detection of variations and their frequencies in the CCR5 gene and its promoter region in Japanese and Okinawan population by NGS analysis using pooled DNA. T. Kaname1, K. Yanagi1, M. Higa1, S. Song1, S. K. Naritomi1, 2) Dept Med Gen, Univ Ryukyus, Nishihara, Japan; 2) Roche Diagnostics K.K., Tokyo, Japan.

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

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

Allelic frequency estimated by the NGS analysis using pooled DNAs almost correlated the real frequency calculated by the individual analysis. Finally, eight novel variations including non-synonymous substitution were found in 260 individuals. For sensitivity to low allelic frequency, 0.5% of allele in the population could be detected at least.

580F
Functional characterization of rare variants in human Dopamine receptor D4 gene by genotype - phenotype correlations. A. Micheal Raj1, N. Atlana2, Md. Jafarulla2, L. Narayanan3, A. Chattopadhyay4, B.K. Thelma5. 1) Department of Genetics, University of Delhi South Campus, New Delhi, Delhi, India; 2) Bioinformatics Center, Sri Venkateswara College, New Delhi, India; 3) Membrane and Receptor Biology Group, Centre for Cellular & Molecular Biology, Hyderabad, India.

Objective: Rapid population growth and weak purifying selection in human populations have been suggested to be causal for the origin of rare variants, many of which could be deleterious and have a major effect of relevance to disease risk. Several rare variants have been reported in dopamine receptor D4 (DRD4) gene, belonging to the GPCR family and associated with many neuropsychiatric conditions. DRD4 is also an important drug target in antipsychotic medication. Therefore, the aim of this study was to identify rare variants of therapeutic relevance in DRD4 using a combination of in silico and in vitro tools. Methods: Cells stably expressing the cDNAs of four target non-synonymous coding rare variants of DRD4 (rs1800443, V194G; rs4991150, 237L; rs3898692, 284G) were generated using lentiviral vectors and tested for functional effects in human neuroblastoma cells. Results: Of the four variants, A281P and S284G ensured to be functionally similar to wild type (WT). V194G variant protein was not able to inhibit forskolin stimulated adenylylate cyclase activity and also failed to phosphorylate the extra cellular receptor kinase due to its insensitive nature towards both dopamine and quinpirole. Further, ligand binding studies showed significant reduction in binding affinity (KD = 2.16 nM; p<0.001) and total number of binding sites (~66%; p<0.001) compared to WT. Potency of dopamine and quinpirole reduced to ~ 6-fold (p<0.01) respectively with V194G when compared to WT protein and ligand binding studies showed reduction only in total number of binding sites (~40%; p<0.01). Ligand docking studies with these two variants revealed that binding of both dopamine (agonist) and spiperone (antagonist) with V194G variant seems to be structurally similar to the WT protein while variant R237L was structurally similar to WT. Conclusion: V194G (at V194G) and arginine (at 237L) residues seem to be important for the activity of DRD4. V194G variant seems to be structurally altered and thus rendered non-functional while R237L seems to be functionally active but with altered expression level. This novel finding is of potential value for lead molecule development for dopamine dysregulated disorders.

Rare genetic variants that severely disrupt protein-coding genes are known as loss-of-function (LoF) variants, and they are of considerable scientific and clinical interest, due to their presumed high probability of being deleterious and having a causal role in severe Mendelian disorders. A previous study surveying exonic variation in imputed low-coverage data and seven high-coverage exomes has estimated that a typical human genome contains approximately 100 LoF variants, though this number is inexact due to incomplete variant discovery. We replicated this study using clinically generated, high-coverage (~100X mean coverage) exome data from over 200 individuals. In addition, we surveyed the frequency of individuals that are heterozygous or homozygous for LoF variants (nonsense mutations, splice site disrupting SNVs, frameshift indels, or larger deletions) in each gene, with further analysis on 1088 candidate ‘essential’ genes: orthologs of mouse (Mus musculus) genes with lethal knockout phenotypes. Only high-quality variants in essential genes were biased in their frequency of less than 1% based on publicly available exome data were considered. An individual genome contains ~280 candidate high quality LoF variants after initial filtering, and 1.4% of genes harbor LoF variants, on average. The proportion of genes mutated is not significantly different for essential versus non-essential genes, but the properties of LoF variants in the two categories differ. Nonsense mutations in essential genes are biased in their effectiveness even amongst proteins in the same complex. A new characterization by independent assessors. A long-term goal for CAGI is to improve the accuracy of phenotype and disease predictions in clinical settings.

583F Identification of nine mutations in the COL1A1 gene in Czech patients with osteogenesis imperfecta. L. Sorvová, I. Falkowská, V. Van Hul, G. Mortier, I. Makí, I. Mazura. 1) Charles University in Prague, 1st Medical Faculty, Prague, Czech Republic; 2) Antwerp University and University Hospital, Centre for Medical Genetics, Antwerp, Belgium; 3) Ambulance Centre for Defects of Locomotor Apparatus, Prague, Czech Republic.

Type I collagen is an abundant protein in connective tissues, especially in bone. The type I collagen molecule is a heterotrimer composed of two copies of the alpha 1(I) procollagen chain (encoded by the COL1A1 gene) and one copy of the alpha 2(I) procollagen chain (encoded by the COL1A2 gene). With this study we aimed to identify the genetic defect in a series of Czech patients with osteogenesis imperfecta (OI) (types IA, II/III, III, IVA and IVB). We elected to restrict the study to the analysis of the COL1A1 gene since 60% of OI mutations are present in this gene. We used gDNA isolated from whole peripheral blood. This research is the first study of the whole coding sequence of the COL1A1 gene of Czech patients affected by OI. The obtained genomic data were analysed using the Ensembl database. Unmapped RNA data variations in the coding and intronic consensus sequences were checked using UCSC Genome Browser (algorithms SpliceSiteFinder, MaxEntScorer, NNSPLICE, GeneSplicer and Human Splicing Finder). We were able to identify the causal mutation in eight patients diagnosed with OI. In three cases with OI IA, we identified nonsense mutations p.Tyr477X at exon 2 (light blue sclera, mild bone deformities, joint hyperlaxity), p.Arg131X at exon 5 (shortened upper body segment, slightly barrel chest) and p.Glu1341X at exon 50 (blue sclera, light bone deformities, barrel chest, higher fractures frequency). In two patients we observed heterozygosity for a missense mutation p.Cys651Phe at exon 2 (OI type III - bone deformities, osteoporosis, barrel chest) and p.Pro1186Ala at exon 48 (OI type IA - blue sclera, mild bone deformities, lower BMD). In the latter patient also a heterozygous nonsense mutation p.Arg415X was found at exon 19. Changes of noncoding sequences were identified in three cases: c.1057-1G>T (intron 16, OI type IAI/IVA - blue sclera, hearing loss, bone deformities, osteoporosis), c.1300-1G>A (intron 19, unclassified OI type, clinical data currently not completed), c.1353+3C>G (intron 20, OI type IA - blue sclera, face asymmetry, joint hyperlaxity, bone deformities, osteoporosis).

584W Findings from the third Critical Assessment of Genome Interpretation, CAGI 2013, a community experiment to evaluate phenotype prediction. S.E. Brenner1, D. Barsky2, J. Moul3, CAGI Participants. 1) University of California, Berkeley, CA; 2) IBBR, University of Maryland, Rockville, MD.

The Critical Assessment of Genome Interpretation (CAGI, ‘k-j’ is a community experiment to objectively assess computational methods for predicting the phenotypic impacts of genomic variation. In the experiment, participants were provided genomic data and made predictions of results. These predictions are evaluated against experimental characterizations by independent assessors. A long-term goal for CAGI is to improve the accuracy of phenotype and disease predictions in clinical settings.

The third CAGI experiment (2012 - 2013) consisted of ten diverse challenges. At time of abstract submission, assessment was underway and results will be known in July 2013. CAGI deliberately extends challenges from previous years, with the continuity allowing measurement of progress. For example, in the second CAGI, in a challenge to predict Crohn’s disease from exomes, one group was able to identify 80% of affected individuals before the first false positive healthy person. In the third CAGI experiment, we targeted genes associated with joint hyperlaxity, and performed remarkably well, with one group achieving a ROC AUC of 0.94 in initial assessment. Another expanded challenge involved using Personal Genome Project genome data to predict phenotypes and match health records; this year, several groups were able to successfully map some genomes. In the expanded challenge to predict benign versus deleterious variants in DNA double-strand break repair MRN genes-Rad50 (from last year), Mre11, and Nbs1–as determined by those that appear in a breast cancer case versus healthy control, predictions show how methods differ sharply in their effectiveness even amongst proteins in the same complex. A new challenge this year was to use exomes from families with lipid metabolism disorders, Familial Combined Hyperlipidemia (FCH) and Hypoalphalipoproteinemia (HA), to predict lipid profiles and a causative variant. Assessment of this challenge is underway. A twist herein real-world data differed sharply from theoretical models. Other CAGI challenges include predicting: cancer impact of BRCA variants; splicing impact of p53 gene variants; and cell proliferation impact of AKT variants. Complementary information about CAGI may be found at http://genominterpretation.org.
585T Whole genome sequencing of rhesus macaques reveals substantial functional variation and justifies a ‘reverse genetics’ approach to identify new models of human disease. M. Raveendran1, D. Rio Deiros2, C.L. Fawcett3, Z. Johnson4, N.H. Kalin5, R.W. Wiseman6, B. Ferguson6, E. Vallender7, S. Kanthasamy8, H. Doddapaneni9, S. Jhangiani10, D.M. Muzny11, R.A. Gibbs12, J. Rogers12,13,14,15. 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Yerkes National Primate Research Center, Emory University, Atlanta, GA; 4) HealthEmotions Research Institute and Dept. of Psychiatry, University of Wisconsin, Madison, WI; 5) Wisconsin National Primate Research Center, Madison, WI; 6) Oregon National Primate Research Center, Oregon Health & Science Univ., Beaverton, OR; 7) New England Primate Research Center, Harvard Medical School, Southborough, MA; 8) California National Primate Research Center, University of California-Davis, Davis, CA.

Rhesus macaques (Macaca mulatta) are the most widely used nonhuman primates in biomedical research, serving as models in multiple areas, including infectious disease, neurobiology, endocrinology and metabolic diseases among others. Traditionally, macaque models of human disease were developed through expensive surveys to identify animals with appropriate phenotypes. To explore an alternative strategy, we generated whole genome sequences (25x coverage) for 51 unrelated Indian-origin rhesus macaques from five NIH primate research centers. Identification of potentially functional genetic variation in this species will create novel opportunities to study targeted genetic models of specific disease mechanisms or genetic pathways in animals much more similar to humans than rodents or other mammalian model organisms. Sequence reads (100bp, paired-end) were mapped to the rhesus genome assembly (rhesac2) using BWA, and variation called using SAMtools mpileup. Variant calls were filtered with a minimum q-score of 20, and only variants observed in two or more animals were retained. This identified >2.7 million SNPs (9.5 SNPs per kb) and >4.2 million small insertions and deletions (1.5 indels per kb), more variation than would be expected in a similar sized survey of humans. Variants were annotated using ENSEMBL to identify those with possible functional effects. We found 48,485 non-synonymous variants will affect the splicing process; however, the ability of algorithms to predict expression of full-length, truncated or no protein is untested. This is a critical issue as the nature and relative abundance of synthesized protein is generally the major determinant of the clinical phenotype caused by a mutation. To address this issue, we developed a novel expression minigene (EMG) system containing flanking intron sequence (~200bp) from multiple adjacent exons cloned into a full-length cDNA in a mammalian expression vector. Presence of the CMV promoter and EMG enables RNA splicing and translation to be tested concurrently. As a proof of concept, two EMGs including introns 11 and 14 to 18 of the CFTR gene were created. Splicing was analyzed by RT-PCR followed by Sanger sequencing and protein translation by Western blot in two cell lines: Human Embryonic Kidney (HEK) 293 cells and Monkey Kidney (hek293) cells. Splicing of the normal pattern of RNA splicing and robust protein translation was confirmed for the two wild-type EMGs. Five mutations associated with cystic fibrosis (c.1585-1G>A, c.1585-2A>G, c.1585-3T>G, c.1585-8G>A, c.1585-9T>A) were introduced into intron 11, three mutations (c.2657+3delG, c.2657+5G>A, c.2657+2_2657+3insA) into intron 16, one in intron 18 (c.2988+1G>A) and one in exon 18 (c.2988G>A). Loss of protein synthesis was observed for 9 of the 10 mutations due to nonsense mediated mRNA decay caused by a frameshift or an in-frame premature stop codon introduced by aberrant splicing. One mutation (c.2657+2_2657+3insA) generated fully processed CFTR protein. We compared our results with two currently available algorithms (ASSEDA and HSF) used to predict the transcript isoforms resulting from these mutations. Overall, HSF correctly predicted the aberrant transcript isoforms for half of the mutations while ASSEDA correctly predicted the normal splicing patterns for 2 of the 10 mutations. Neither of these programs predicted whether protein would be synthesized. Despite the limitations of current algorithms, experimental data generated by EMGs should enable the generation of algorithms that accurately predict the effect of splicing mutations upon protein production.

586F Expression Minigenes Reveal the Limitations of Algorithms Predicting the Consequences of Putative Splice Site Mutations. N. Sharma1, P. Sosnowski1, C. Frangal2, C. Current3, A. Franco4, C. Hsiung5, L.B. Gottschalk4, K.R. siklosi1, M. Amaral4, R. Karchin2, G.R. Cutting1,2,3. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD; 4) University of Lisbon, Faculty of Sciences, BioFIG - Centre for Biodiversity, Functional and Integrative Genomics, Lisboa, Portugal; 5) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Splice sites flanking each exon direct the generation of mature mRNA from heteronuclear RNA. Algorithms that utilize sequence conservation in splice sites provide reasonably accurate predictions as to whether DNA variants will affect the splicing process; however, the ability of algorithms to predict expression of full-length, truncated or no protein is untested. This is a critical issue as the nature and relative abundance of synthesized protein is generally the major determinant of the clinical phenotype caused by a mutation. To address this issue, we developed a novel expression minigene (EMG) system containing flanking intron sequence (~200bp) from multiple adjacent exons cloned into a full-length cDNA in a mammalian expression vector. Presence of the CMV promoter and EMG enables RNA splicing and translation to be tested concurrently. As a proof of concept, two EMGs including introns 11 and 14 to 18 of the CFTR gene were created. Splicing was analyzed by RT-PCR followed by Sanger sequencing and protein translation by Western blot in two cell lines: Human Embryonic Kidney (HEK) 293 cells and Monkey Kidney (hek293) cells. Splicing of the normal pattern of RNA splicing and robust protein translation was confirmed for the two wild-type EMGs. Five mutations associated with cystic fibrosis (c.1585-1G>A, c.1585-2A>G, c.1585-3T>G, c.1585-8G>A, c.1585-9T>A) were introduced into intron 11, three mutations (c.2657+3delG, c.2657+5G>A, c.2657+2_2657+3insA) into intron 16, one in intron 18 (c.2988+1G>A) and one in exon 18 (c.2988G>A). Loss of protein synthesis was observed for 9 of the 10 mutations due to nonsense mediated mRNA decay caused by a frameshift or an in-frame premature stop codon introduced by aberrant splicing. One mutation (c.2657+2_2657+3insA) generated fully processed CFTR protein. We compared our results with two currently available algorithms (ASSEDA and HSF) used to predict the transcript isoforms resulting from these mutations. Overall, HSF correctly predicted the aberrant transcript isoforms for half of the mutations while ASSEDA correctly predicted the normal splicing patterns for 2 of the 10 mutations. Neither of these programs predicted whether protein would be synthesized. Despite the limitations of current algorithms, experimental data generated by EMGs should enable the generation of algorithms that accurately predict the effect of splicing mutations upon protein production.

587W Linear Decay of Retrotransposon Antisense Bias across Genes is Contingent upon Tissue Specificity. S. Linker1, D. Hedges1,2. 1) HIHG, University of Miami, Miami, FL; 2) Division of Human Genetics, The Ohio State University, Columbus, Ohio.

Retrotransposons (RTs), which make up approximately half of the human genome, are gaining attention due to new techniques which allow polymorphic RT insertions to be identified from next-generation sequencing data. The ability to move forward and infer the effects of these polymorphisms will require a substantial increase in the understanding of the complex effects that retrotransposons have on a coincident gene. Previous work in this area has highlighted the non-uniform distribution of retrotransposons across the genome leading researchers to propose that there may be local effects that modify their retention in various genomic contexts. Our work furthers this effort by determining predictable trends of retrotransposon accumulation as well as through identifying links between the presence of retrotransposons and gene expression. Our primary findings show that retrotransposons which are in the antisense orientation with respect to a gene, exhibit a linear decay in frequency across the length of the gene. Conversely, retrotransposons in the sense direction maintain a low-level frequency independent of location within a gene. This linear trend of antisense RTs is inversely correlated with exon frequency, which we show has a linear increase in frequency across the length of most genes. Interestingly, this correlation is dependent upon the gene type (ie: housekeeping versus tissue-specific). Moreover, we have found that transcription factor binding motifs which are contained within RTs also exhibit a pattern of accumulation in certain regions of the gene. We used retrotransposons which are polymorphic in the human genome to determine the potential for effect on expression of these inserts. Indeed we identified a subset of polymorphic RTs that were significantly correlated with altered gene expression after correction with FDR. These findings begin to map out predictable variables for the retention of RTs in the genome, and may aid researchers in inferring the effects of novel polymorphic RTs on local gene expression.
588T  
Nicotine causes genome-wide microsatellite instability in normal epithelial cells. J. Bavara, L. Molver, T. Hongseok, H. Garner. Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA.

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

589F  
Probing genes for hyperphagia in rare obesity-related disorders. M.G. Butler1, J.D. Marshall2, J. Rethemeyer1, K. Wang3, A.M. Manzardo4, 1) Psychiatry & Behavioral Sciences, University of Kansas Medical Center, Kansas City, KS; 2) The Jackson Laboratory, Bar Harbor, ME.

Hyperphagia and obesity are key features of several rare genetic obesity-related disorders including Prader-Willi syndrome (PWS) and Alström syndrome (ALMS). A better understanding of genetic causes through the study of obesity-related disorders should provide a more comprehensive picture of mechanisms that control food intake and energy balance related to the development of obesity. Comparison of gene expression patterns associated with rare mutations should provide insights into commonly disturbed gene pathways involved in appetite control and body weight regulation with application to obesity in the general population. We compared coding and non-coding gene expression in PWS, ALMS and nonsyndromic obesity relative to non-obese adult male controls using both lymphoblasts and brain tissue coding gene expression in PWS, ALMS and nonsyndromic obesity relative to non-obese adult male controls using both lymphoblasts and brain tissue. The analysis of this heterogeneity has become a focus of interest in various fields of biology, especially in stem cell research. The main obstacles to analysis of gene expression at the single-cell level are the low amount of starting material and the low abundance of many transcripts of interest. This requires a high level of confidence in results obtained from unique samples, making it difficult to be done accurately by traditional quantification methods such as qPCR. Droplet digital PCR (ddPCR™) provides absolute quantification of individual molecules with high precision, and without the requirement for standard curves or pre-amplification steps. Using Bio-Rad’s QX100 ddPCR system, we developed a method that measures single-cell gene expression in multiplexed assays with high sensitivity and reproducibility, thus enabling us to simultaneously analyze expression of different targets in cDNA from the same cell. In order to perform this, we carefully evaluated various cell lysis and cDNA synthesis methods and developed a protocol with flow-sorted Jurkat cells that is fully compatible with ddPCR, easy to use and capable of analyzing gene expression in single cells without pre-amplification. We show that high, medium and low abundance transcripts (< 20 copies/cell) can be reproducibly measured with ddPCR. We further used this protocol to measure cell cycle-specific genes and revealed distinctive gene expression patterns in populations of single cells. This method allows us to easily and quickly measure the expression of multiple genes of interest in single cells, minimizing the stochastic effect of sampling and empowering us to accurately and sensitively detect and quantify low-expressing genes in single cells.

591T  
Individual variation in the rate of retrotransposition in iPSC cells and its effect on genomic instability with regard to medical utility. T.T. Doucet1, C. Smith1,2, A. Ewing3, K. Burns1,2, L. Cheng1,4, H.H. Kazazian1, 1) Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD USA; 2) Pre-doctoral training program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA; 3) Center for Biomolecular Science and Engineering, University of California at Santa Cruz, Santa Cruz, California 95064, USA; 4) Stem Cell Program, Institute for Cell Engineering, Division of Hematology, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 5) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

Genomic instability can be caused by repetitive sequences in the human genome (retrotransposons) that can copy and paste themselves into new sites. We seek to understand how retrotransposition plays a role in genomic instability during the reprogramming process as well as the differentiation process of iPS cells as a medical treatment, their genomes must be stable. We are evaluating iPSC and parental cell lines, the cells from which the iPSC were derived, from three individuals utilizing next generation sequencing, RT qPCR, and immunohistochemistry. To ensure we find all the putative somatic insertions in the iPSC lines, we have used a LINE-1 enrichment technique, L1-seq, prior to next generation sequencing. To validate putative somatic insertions in the iPSC cells, we will use site specific PCR, a LINE-1 specific primer and a genome specific primer derived from an artificial chromosome and Santorin. The presence of LINE-1 mRNA and LINE-1 associated proteins such as ORF1p and ORF2p are positive indicators of retrotransposon activity. Using the RNA isolated from both the parental lines and the iPSC lines we have demonstrated that the expression of LINE-1 RNA in the IPS lines evaluated compared to the expression in the parental lines. We have also seen ORF1 protein expression in the IPS cells, which is not present in the parental cells. To evaluate the differences in the rate of retrotransposition between individuals’ IPS lines, for IPS cells to be utilized as a medical treatment, their genomes must be stable. We are evaluatingips and parental cell lines, the cells from which the IPS cells were derived, from three individuals utilizing next generation sequencing, RT qPCR, and immunohistochemistry. To ensure we find all the putative somatic insertions in the IPS cells, we have used a LINE-1 enrichment technique, L1-seq, prior to next generation sequencing. To validate putative somatic insertions in the IPS cells, we will use site specific PCR, a LINE-1 specific primer and a genome specific primer derived from an artificial chromosome and Santorin. The presence of LINE-1 RNA in the IPS lines evaluated compared to the expression in the parental lines. We have also seen ORF1 protein expression in the IPS cells, which is not present in the parental cells. To evaluate the differences in the rate of retrotransposition between individuals’ IPS lines, for IPS cells to be utilized as a medical treatment, their genomes must be stable. We are evaluatingips and parental cell lines, the cells from which the IPS cells were derived, from three individuals utilizing next generation sequencing, RT qPCR, and immunohistochemistry. To ensure we find all the putative somatic insertions in the IPS cells, we have used a LINE-1 enrichment technique, L1-seq, prior to next generation sequencing. To validate putative somatic insertions in the IPS cells, we will use site specific PCR, a LINE-1 specific primer and a genome specific primer derived from an artificial chromosome and Santorin. The presence of LINE-1 RNA in the IPS lines evaluated compared to the expression in the parental lines. We have also seen ORF1 protein expression in the IPS cells, which is not present in the parental cells. To evaluate the differences in the rate of retrotransposition between individuals’ IPS lines, for IPS cells to be utilized as a medical treatment, their genomes must be stable. We are evaluatingips and parental cell lines, the cells from which the IPS cells were derived, from three individuals utilizing next generation sequencing, RT qPCR, and immunohistochemistry. To ensure we find all the putative somatic insertions in the IPS cells, we have used a LINE-1 enrichment technique, L1-seq, prior to next generation sequencing. To validate putative somatic insertions in the IPS cells, we will use site specific PCR, a LINE-1 specific primer and a genome specific primer derived from an artificial chromosome and Santorin. The presence of LINE-1 RNA in the IPS lines evaluated compared to the expression in the parental lines. We have also seen ORF1 protein expression in the IPS cells, which is not present in the parental cells. To evaluate the differences in the rate of retrotransposition between individuals’ IPS lines, for IPS cells to be utilized as a medical treatment, their genomes must be stable. We are evaluatingips and parental cell lines, the cells from which the IPS cells were derived, from three individuals utilizing next generation sequencing, RT qPCR, and immunohistochemistry. To ensure we find all the putative somatic insertions in the IPS cells, we have used a LINE-1 enrichment technique, L1-seq, prior to next generation sequencing. To validate putative somatic insertions in the IPS cells, we will use site specific PCR, a LINE-1 specific primer and a genome specific primer derived from an artificial chromosome and Santorin. The presence of LINE-1 RNA in the IPS lines evaluated compared to the expression in the parental lines. We have also seen ORF1 protein expression in the IPS cells, which is not present in the parental cells.
592F...IqB complexes and kinetics in non-secretors, ...expression pattern of NF-IqB monitored by immunoblot analysis of whole cell lysates...and its kinetics were examined by immunoblot analysis in whole cell lysates. Expression of DcR3 and NF-IqB were examined by immunoblot analysis in whole cell lysates...in TNFRSF6B (A allele) exhibit differential pattern of DcR3 expression and were evaluated by western blots...B complexes and kinetics in non-secretors, ...Expression pattern of NF-IqB monitored by immunoblot analysis of whole cell lysates...and its kinetics were examined by immunoblot analysis in whole cell lysates. Expression of DcR3 and NF-IqB were examined by immunoblot analysis in whole cell lysates...DcR3 knockdown was performed using specific DcR3 siRNA. Cell proliferation and cell death was measured by MTT assay. Caspase8, caspase9, caspase10, Bcl2 expression were determined by western blots...Results: EBV transformed cell lines derived from IBD patients harboring risk variants in TNFRSF6B (A allele) exhibit differential pattern of DcR3 expression and NF-xB activation kinetics in comparison with wild type. siRNA mediated knockdown post 24hrs of nucleofection results in decreased DcR3 expression, increased cell death and decreased cell proliferation, effects that were knockdown post 24hrs of nucleofection results in decreased DcR3 expres-sion, increased cell death and decreased cell proliferation, effects that were...in TNFRSF6B and WWOX, respectively, and FRAG instability drives amplification of the MET oncogene. Several hypotheses have been proposed to explain the mechanism leading to breaks at fragile sites. One hypothesis proposes that a lack of origins in fragile site regions leads to incomplete replication when replication is slowed, causing breaks in un-replicated regions during anaphase. A second hypothesis proposes that fragile sites contain AT-rich sequences that are considered to be flexible and prone to forming secondary structures during replication, which can cause further replication stalling and replication fork collapse. We are evaluating these hypotheses using yeast artificial chromosomes (YACs) containing inserts of human DNA from fragile sites FRA3B, FRA4H, and FRA8B. The YAC insert sizes are 1.3 Mb, 730 Kb, and 362 Kb, respectively. We have placed yeast under replication stress, caused by low levels of polymerase alpha, to induce fragile site breaks. We have mapped 30 break locations in each YAC to a 6 Kb resolution. In all three YACs, there is a hotspot for breaks near the distal end of the human DNA fragile site insert. We found that 60% of FRA3B breaks are clustered within a 15 Kb region near the centromeric end of FRA3B. Analysis of the different forms of the HIV-1 genome showed that the defect in HIV-1 replication occurred prior to late RT product formation. This inhibition of HIV-1 is dependent on the HIV-1 envelope glycoprotein since viral particles pseu-dotyped with the vesicular stomatitis virus glycoprotein (VSVg) were unaffected. The identification of a novel role for HDFs in early events in the HIV-1 life cycle could suggest potential novel therapeutic approaches for the inhibition of viral transmission and replication...593W...Functional characterization of DrC3 in EBV transformed cell lines from IBD patients of different allelic background and role in disease pathogenesis. R. Pandey1, C. Cardinale1, K. Kachelries2, S.F.A. Grant1,3, R. Baldassano1,3, H. Hakonarson1,3,1, 1) Centre for Applied Genomics, Children’s Hospital Of Philadelphia, Philadelphia, PA; 2) Division of Gastroenterology, Hepatology and Nutrition, Children’s Hospital of Philadelphia; 3) Department of Pediatrics, Perelman School of Medicine at the University of Philadelphia, PA. Aim: DrC3, a receptor of the TNFR superfamily is a soluble receptor for FasL and plays significant role in immune suppression and tumor progression by neutralizing the Fasl mediated apoptosis signal. Here we investigate the possible immuno-modulation mediated by DrC3 and NF-xB in EBV transformed control and patient derived cell lines with and without risk variants in the TNF Receptor Superfamily 68 gene, TNFRSF6B captured by the tagging SNP, rs2315008. Methods: Expression of DrC3 and its kinetics were examined by immunoblot analysis in whole cell lysates from EBV transformed control and patient derived cell lines of different genotype background for rs2315008 (AA, AT, TT). DrC3 induced rapid activation of nuclear factor xB (NF-xB) monitored by immunoblot analysis of IKBx. Expression pattern of NF-xB complexes and kinetics in non-secretors, control and patient-derived EBV transformed cell lines was examined by immunoblot analysis in whole cell lysates, cytoplasmic and nuclear extracts. DrC3 knockdown was performed using specific DrC3 siRNA. Cell proliferation and cell death was measured by MTT assay. Caspase8, caspase9, caspase10, Bcl2 expression were determined by western blots. Results: EBV transformed cell lines derived from IBD patients harboring risk variants in TNFRSF6B (A allele) exhibit differential pattern of DcR3 expression and NF-xB activation kinetics in comparison with wild type. siRNA mediated knockdown post 24hrs of nucleofection results in decreased DcR3 expression, increased cell death and decreased cell proliferation, effects that were...in TNFRSF6B and WWOX, respectively, and FRAG7 instability drives amplification of the MET oncogene. Several hypotheses have been proposed to explain the mechanism leading to breaks at fragile sites. One hypothesis proposes that a lack of origins in fragile site regions leads to incomplete replication when replication is slowed, causing breaks in un-replicated regions during anaphase. A second hypothesis proposes that fragile sites contain AT-rich sequences that are considered to be flexible and prone to forming secondary structures during replication, which can cause further replication stalling and replication fork collapse. We are evaluating these hypotheses using yeast artificial chromosomes (YACs) containing inserts of human DNA from fragile sites FRA3B, FRA4H, and FRA8B. The YAC insert sizes are 1.3 Mb, 730 Kb, and 362 Kb, respectively. We have placed yeast under replication stress, caused by low levels of polymerase alpha, to induce fragile site breaks. We have mapped 30 break locations in each YAC to a 6 Kb resolution. In all three YACs, there is a hotspot for breaks near the distal end of the human DNA fragile site insert. We found that 60% of FRA3B breaks are clustered within a 15 Kb region near the centromeric end of FRA3B. Analysis of the different forms of the HIV-1 genome showed that the defect in HIV-1 replication occurred prior to late RT product formation. This inhibition of HIV-1 is dependent on the HIV-1 envelope glycoprotein since viral particles pseudotyped with the vesicular stomatitis virus glycoprotein (VSVg) were unaffected. The identification of a novel role for HDFs in early events in the HIV-1 life cycle could suggest potential novel therapeutic approaches for the inhibition of viral transmission and replication.
596W
Insulin-Induced Kinase Signaling Contributes to Individual Differences in Response to Insulin. I.X. Wang1, V.G. Cheung2,3, 1) Howard Hughes Medical Institute; 2) Genetics/Pediatrics, Univ Pennsylvania, Philadelphia, PA.

Individual differences in sensitivity to insulin underlie diseases such as diabetes and influence cell growth and cancer susceptibility. Insulin mediates cellular functions by triggering signal transduction pathways that lead to subsequent changes in gene expression. In this study, we focused on individual variation in signal transduction. Even though signaling is the first step in cellular response to nutritional demands, individual variability in insulin-induced activation of signaling pathways is poorly understood. To address this, we exposed primary skin cells from normal individuals to insulin and measured phosphorylation of kinases. We found extensive individual variation in insulin-induced activation of key signaling factors, including ERK whose induction differs by more than 20-fold among our subjects. By genetic analysis, we identified DNA variants that influence signaling response and downstream changes in gene expression and cell proliferation. The effect of variation in kinase activation on gene expression is substantial; for example, differences in ERK activation contribute to almost 14% of individual variation in insulin-induced changes in gene expression. To better understand the relationship between the signaling proteins and target genes, we took advantage of the variability and constructed networks which identify the connections within the signaling pathways and the extensive crosstalk between signaling regulators and their target genes. In this presentation, I will describe our results that demonstrate how signal transduction is an important contributor to insulin sensitivity, therefore offer kinase modulators as promising therapeutics for diseases characterized by insulin resistance.

597T

According to the aging theory presented by Harman, the production of free radicals rises with age and plays a key role in the degenerative processes of senescence. This oxidative stress increase could be the origin of cellular molecule damage. Particularly, the highest levels of oxidative stress are generated in the mitochondria due to the Electron Transport Chain (ETC). This oxidative stress increase induces accumulation of nonrepaired lesions in mitochondrial DNA (mtDNA). There are some studies that point out the relationship between mtDNA mutations and age in different tissues. These studies are potentially interesting for forensic identification because they can help to improve the estimation of age at death. Since teeth are the hardest tissue of human body, and one of the most abundant types of biological remains available in forensic cases, the present study aims to evaluate the mutations in mtDNA from dentin and pulp and their relation with the age. 30 healthy erupted third molars were extracted for valid clinical reasons from individuals from the Northwestern region of Spain aged 20 to 70. DNA was isolated from the dentin and pulp of each molar. We used specific restriction enzymes for region 2 (HVI2) of the mitochondrial D-loop to assess the mutations in each type of tissue by PCR. We found a decrease in the amplification of this region with age in dentin and variations may be involved in heat-shock or stress related responses and we also identified a novel mutation. This network approach can only interrogate individual proteins and it fails to uncover any changes in DNA binding levels between the two conditions. However, this proof-of-concept in yeast is the first evaluation of ATHENA for epigenetic data. Our model identified important protein factors that could differentiate the two conditions with 80–90% accuracy based on their genome-wide binding levels. Our top models consisted of approximately 10 protein factors selected if they can distinguish based on how well they can predict both 70% groups under normal condition or genes under heat shock condition when interacting with other protein factors. We replicated several protein factors known to be involved in heat-shock or stress related responses and we also identified a few new factors, especially for the gene term methylated CpG islands. These models showed that the protein factors exhibited non-additive relationships, indicating the presence of interactions among the protein factors. We also uncovered different regulation networks at the gene promoter regions and heat-shock regions, which showed that different sets of factors were affected by heat shock at distinctive biological locations. This machine learning approach demonstrates the ability to uncover interactions in protein binding data with high accuracy and it has the potential to be extended to any type of genomics data.

598F
Genome-wide analysis identifies heat shock induced gene and chromatin regulatory protein network in Saccharomyces cerevisiae. R. Li5, M.D. Ritchie1,5, 1) Bioinformatics and Genomics, The Pennsylvania State University, State College, PA; 2) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, State College, PA.

In Saccharomyces cerevisiae, hundreds of different proteins are involved in transcription regulation, yet much of the regulation mechanisms remain largely unknown. Previous studies have gained insights into transcription regulation from studying the protein factor binding responses induced by heat shock. We have limited in the capacity to comprehensively study of almost all DNA binding proteins, chromatin regulators, general transcription proteins, and elongation regulators in Saccharomyces generated ~200 proteins’ genome-wide binding levels using ChIP-chip in normal and heat shock conditions. Basic statistical approaches have been used to identify important proteins that exhibit the highest fold changes in DNA binding levels between the two conditions. However, this approach can only interrogate individual proteins and it fails to uncover any potential interactions among the proteins involved in gene regulation. Thus we applied ATHENA (the Analysis Tool for Heritable and Environmental Network Associations), which is a software package for modeling the underlying interaction relationships between protein factors using a machine-learning technique, grammatical evolutionary neural networks (GENN). This proof-of-concept in yeast is the first evaluation of ATHENA for epigenetic data. Our model identified important protein factors that could differentiate the two conditions with 80–95% accuracy based on their genome-wide binding levels. Our top models consisted of approximately 10 protein factors selected if they can distinguish based on how well they can predict both 70% groups under normal condition or genes under heat shock condition when interacting with other protein factors. We replicated several protein factors known to be involved in heat-shock or stress related responses and we also identified a few new factors, especially for the gene term methylated CpG islands. These models showed that the protein factors exhibited non-additive relationships, indicating the presence of interactions among the protein factors. We also uncovered different regulation networks at the gene promoter regions and heat-shock regions, which showed that different sets of factors were affected by heat shock at distinctive biological locations. This machine learning approach demonstrates the ability to uncover interactions in protein binding data with high accuracy and it has the potential to be extended to any type of genomics data.

599W
Human neuropathy target esterase rescues SWS Drosophila neurodegeneration. N.D. Hein1, A.K. Sujkowski2, A.M. Taylor3, R.J. Tabano4, A. Harrell1, S. Rainier1, R.J. Wessell5, J.K. Fink1,2, 1) Neurology, University of Michigan, Ann Arbor, MI; 2) Internal Medicine, University of Michigan, Ann Arbor, MI; 3) Geriatric Research Education and Clinical Center, Ann Arbor Veterans Affairs Medical Center.

Human neuropathy target esterase (NTE) is a membrane phospholipase A2 with Camp-dependent protein kinase (PKA) regulatory activity. NTE inhibition by organophosphorous (OP) compounds leads to chronic OP-induced delayed neuropathy. NTE mutations cause autosomal recessive motor neuron disease (SWS-MND), a hereditary genetic disorder (RH9). Drosophila Swiss cheese (SWS) shares 39% protein sequence identity with human NTE (hNTE). Drosophila SWS mutants exhibit progressive locomotor impairment, early lethality, and vacuolar degeneration and glial hyperplasia in brain. SWS5 has mutation G648R. SWS Drosophila neurodegeneration can be rescued by wildtype sws or wildtype murine NTE transgenes. We examined the ability of wildtype hNTE to ameliorate the SWS neurodegenerative phenotype. G648R transgenic for RU486-inducible elav-GAL4 (generated by A. Sujkowski) were crossed with transgenic UAS-NTE flies; and progeny crossed with SWS5 flies (generously provided by Dr. Doris Kretzchmar, Oregon Science & Health Univ.) and control flies to create flies with RU486-inducible, neuron-specific hNTE expression. Flies were maintained at 29°C and locomotor ability assessed at day 10 in a run-to-exhaustion protocol using iterative negative geotaxis (Tinkerhess et al, 2012). NTE protein (western blot), PKA activity, and NTE activity (phenyl valerate substrate) were assessed using published methods. Western blot analysis confirmed the presence of hNTE protein in flies expressing wildtype hNTE (RU486-induced SWS5, elav-GAL4;NTE and Berlin K5;elav-GAL4;NTE control flies). Locomotor activity of SWS5 flies was approx. half that of wildtype flies. SWS5 flies expressing wildtype NTE (RU486-induced SWS5/;elav-GAL4;NTE) had locomotor activity very similar to control flies. Histopathology studies of SWS5, SWS5/;elav-GAL4;NTE and control flies is in progress.

Conclusion. These studies demonstrate the capacity of hNTE to functionally compensate for the SWS5 Drosophila neurodegenerative phenotype. Together with the previously reported correction of SWS mutant phenotype by mNTE, these studies underscore the functional homology between mammalian NTE and SWS proteins. This in vivo assay of hNTE functional activity can be used to evaluate the functional significance of identified NTE variants, including those in NTE-MND subjects.
600T Microbiomic profiles and clinicopathologic outcome markers in oropharyngeal cancers. P. Funkhouser, G. Bebek, K. Bennett, N. Fowler, B. Burkey, C. Eng, N. Fouquet, C. Wrestling, K. Kellar, D. Cleveland Clinic, Cleveland, OH; 2) Department of Genetics and CARE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH; 3) Case Center for Proteomics and Bioinformatics Case Western Reserve University, Cleveland, OH; 4) Departments of Head and Neck Institute, Cleveland Clinic, Cleveland, OH.

Recent studies of the human microbiome suggest that non-human genetic material from the spectrum of commensal microorganisms which inhabit the human body, otherwise known as the human microbiome, provide a large contribution to host health and disease, strongly in particular for the digestive tract. In the aerodigestive tract, the microbiome shapes the immune system and is largely influenced by environmental factors, host genetics also play an important role. The importance of understanding the effects of host genetics on the microbiome, and how host-microbe interactions can influence an individual's susceptibility to disease.


1) Ecolgy & Evolution, Stony Brook University, Stony Brook, NY; 2) Departments of Human Genetics, and Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 3) Department of Genetics, Stanford University, Stanford, CA 94305; 4) The J. David Gladstone Institutes, University of California, San Francisco, San Francisco CA, 94158; 5) Department of Structural Biology, Stanford University, Stanford CA 94305; 6) University of Cambridge, Cambridge CB2 1ST, UK; 7) Agilent Technologies, Genomic Division, Cedar Creek, TX 78612; 8) Institute for Human Genetics, and the Departments of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco CA, 94143; 9) Department of Biological Sciences, Stanford University, Stanford, California 94305.

Targeted capture of genomic regions reduces sequencing cost while generating higher coverage by allowing biomedical researchers to focus on specific loci of interest, such as exons. Targeted capture also has the potential to facilitate the generation of genomic data from DNA collected via saliva or buccal cells. DNA samples derived from these cell types tend to have a lower human DNA yield, may be degraded from age and/or have contamination from bacterial or other ambient oral flora. However, thousands of samples have been previously collected from these tissue types, and saliva collection has the advantage that it is a non-invasive form of collection, appropriate for a wide variety of research. We demonstrate successful enrichment and sequencing of 15 South African Khoesan exomes with samples initially derived from saliva. The expanded exome dataset enabled us to more accurately map and call the highly polymorphic HLA and KIR loci from exome capture data. Finally, we show that capture of saliva-derived DNA yields sufficient non-human sequences to characterize oral microbiotal communities, including detection of bacteria linked to oral disease (e.g. Prevotella in periodontal disease). Metagenomic profiles obtained from exome-capture results are similar to those obtained from direct sequencing of saliva derived DNA, indicating that metagenomic analysis of saliva derived samples holds promise for future metagenomic studies as a ‘free’ addition to human exome sequencing.


1) Department of Cardiovascular Medicine, Osaka University Graduate School of Medicine, Suita, Osaka, Japan; 2) Department of Genome Informatics, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, Japan; 3) Department of Genetics, Osaka University Medical School, Osaka, Japan.

Infective endocarditis (IE) is still a lethal disease and detecting causative microorganisms is essential for choosing an optimal antibiotic and effective treatment. However, culture-negative IE is frequently observed, especially among cases that included premitral valve with antibiotic pretreatment. Conventional valve culture has disadvantages that it has low sensitivity and can detect only viable and target microorganisms with the culture medium currently available. Recently, the metagenomic approach using next generation sequencing (NGS) technology emerged as a comprehensive method for exploring causative agents of infectious diseases without prior culture. Therefore, we assessed the viability of metagenomic analyses to detect causative microorganisms in resected valves from IE patients who had undergone surgical valve replacement with antibiotic pretreatment, each of which is a patient with culture-positive IE and culture-negative IE, respectively. After the operation of a valve replacement therapy for IE we assessed causative bacteria in the resected valve both by cultivation survey and by metagenomic sequencing analysis. The former case was affected with IE of the native aortic valve. The blood culture at the primary hospital was positive and negative at the time of referral to our hospital with antibiotic pretreatment. After the operation the resected valve culture was positive for the same bacteria species as detected in the previous hospital. By defining this patient as ‘case with culture-positive IE’ we applied metagenomic approach to detect bacterial genome fragments with genome DNA extracted from the resected valve. Using BLAST search, the dominant parts in bacteria of IE were identified in pathological specimen, in which Gram-positive cocci were identified in the valve tissue. In conclusion, comprehensive metagenomic approach using NGS could detect causative organisms in culture-negative IE case.
Temporal variation in human gut microbiome composition in the Huttenetes. E.R. Davenport, O. Mizrahi-Man, K. Michelinii, L.B. Barreiro, C. Ober, Y. Gilad. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Sainte Justine Hospital Research Centre Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, QC, Canada.

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


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

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

Psoriasis is an inflammatory disease of skin and joints affecting 0.1–2% of the world’s population. Previous studies have identified crucial immune and epidermal-differentiation genes that are differentially expressed in the lesional skin of psoriatic patients. However, most of these studies assayed only protein-coding genes. The functional importance of long non-coding RNA (lncRNAs), which are known to play a role in human disease, is yet to be determined for psoriasis. In this study, we applied a computational approach and stringent filtering to predict novel IncRNAs in our RNAseq data comprised of 92 psoriatic and 82 normal skin samples, and we characterized the expression patterns of the identified IncRNAs to infer their biological functions. We identified 7,145 known IncRNAs, and further predicted 319 novel ones, yielding on average genomic density of 2.5 expressed IncRNAs per Mb. Although the proportions of differentially expressed genes for protein-coding and known IncRNAs were similar (18%), over 60% of our novel IncRNAs were differentially expressed. This result suggests the novel IncRNAs identified here have a skin-specific expression pattern and function. Notably, the transcript with the most significant correlation between expression levels and disease severity as assessed by the local psoriasis area and severity index (PASI) is a lncRNA. We used co-expression analysis to infer the biological functions of IncRNAs, and the most enriched functions include immune response (p=2×10⁻⁹) and cytokine interactions (p=6×10⁻¹⁰). Our results also indicate significant enrichment (p=5×10⁻³) of novel IncRNAs in the epidermal differentiation complex on chromosome 1q21. The characterization of novel IncRNAs provided by this study suggests an important role for these transcripts in disease etiology.

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

The orchestration of brain function requires complex gene regulatory networks, which in part, are modulated by microRNAs (miRNAs). These non-coding RNAs associate with Argonaute (Ago) proteins to direct post-transcriptional gene suppression by binding to 3’ untranslated regions (3’-UTRs), and growing evidence suggests that even slight aberrations in miRNA activities may alter synaptic function. To better understand how miRNAs contribute to human-specialized brain processes and neurological phenotypes, identifying their targets is of paramount importance. Here, we address the latter by profiling Ago2:RNA interactions using crosslinking immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) to generate the first transcriptome-wide map of miRNA binding sites in human brain. We uncovered ~7000 stringent Ago2 binding sites which are highly enriched for conserved sequences corresponding to abundant brain miRNAs. This dataset points to functional miRNA:target pairs across more than 3000 genes and represents a valuable resource for accelerating our understanding of miRNA function in the central nervous system. We explored this interactome for clinically-relevant miRNA binding sites and identified numerous disease-associated targets for mir-137, a miRNA implicated in schizophrenia. In addition, we discovered miRNA binding sites overlapping single nucleotide polymorphisms linked to Parkinson and rare neuropsychiatric conditions, including Parkinson’s and Alzheimer’s. These findings provide clues which may facilitate the translation of genetic studies of complex neuropsychiatric diseases into novel or refined pathogenic mechanisms and therapeutics. Overall, this work lays the foundation for translating this methodology to characterize the diverse landscapes of miRNA target interactions throughout sub-anatomical brain structures and across normal and diseased tissues.
De novo discovery of distant regulatory elements by enhancer RNA expression. H. Wu1, A. Nord1, J. Akiyama1, M. Shoukry1, V. Afzal1, E. Rubin1,2, L. Pennacchio1,2, A. Visel1,2. 1) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA; 2) DOE Joint Genome Institute, Walnut Creek, CA.

Enhancer DNA elements can be actively transcribed to produce short and long transcripts, named enhancer RNAs or eRNAs. Recently, several groups showed that expression of eRNAs appears to correlate with enhancer activity in cell lines. However, whether this is true and can be used to identify tissue-specific in vivo enhancers remain unknown. Here we have investigated the expression dynamics of eRNAs in mouse embryonic tissues (E11.5), heart and limb. We show that active enhancers are transcribed in vivo, and there are clear differences in the expression dynamics of eRNAs in different tissues. Furthermore, we show that expression of eRNAs appears to correlate with enhancer activity in cell lines. This indicates that eRNAs may be used as a potential tool to identify enhancers in vivo.

Association of non-coding SNPs with decreased levels of miR-9 and alcoholism. A. Pietrzynowski1,2, Y. Wang3, O. Aneesa, E. Meada, N. Kistilinger1, C. Tejeda1, A. Hoft1. 1) Animal Sciences, Rutgers University, .; 2) Berwick National Laboratory, . Berkeley, CA; 2) DOE Joint Genome Institute, . Lawrence Berkeley National Laboratory, . Berkeley, CA.

Alcoholism has strong genetic underpinnings. The focus on protein-coding genes has only partially revealed the genetic basis of this complex, multifactorial disorder. There is increasing recent evidence that microRNAs (miRNAs) play a key role in alcoholism and other types of addiction. miRNAs are produced by non-coding genes, whose products are small RNA molecules with a powerful role as master regulators of mRNA and protein expression. Many miRNAs can simultaneously control the expression of several genes. Our previous results using rodent models indicated that a particular miRNA, miR-9, is regulated by alcohol and has an essential role in the development of alcohol tolerance to this drug. Here, we seek to determine an association of single nucleotide polymorphisms (SNPs) in the regions regulating miR-9 expression in alcohol using human samples. miR-9 is encoded by three distinct genes, with each gene located within a larger host gene. Thus, six different promoters can control expression of mature miR-9. We used the UCSC Genome Browser to determine the length of each promoter (4 to 6 kb) and perform SNP discovery in all 6 regions. We used 282 alcoholic samples from the Collaborative Studies on Genetics of Alcoholism (COGA) collection, and 255 non-alcoholic controls from the NIMH control collection. Nested PCR and direct sequencing of PCR products were performed on each sample. GWAS-derived ancestry informative markers were used to correct for ancestry. Association, allele frequency, odds ratio, heterozygosity and Hardy-Weinberg equilibrium were determined for all samples. MAPPFinder or Malinator were used to test the effects of SNPs in 10 different pathways. We observed evidence for association between human alcohol use disorder and expression of miR-9 in the inner ear. One example is a miRNA gene that is included in the intron of the gene Tectb, which is expressed in the mouse inner ear and is associated with deafness. The gene produces two mature miRNAs, 5p and 3p, and both are upregulated in human alcoholics relative to controls.

MicroRNA regulation in the inner ear and link to deafness. K. Avraham1, A. Rudnicki1, K. Ushakov2, O. Isakso3, N. Shomron4, 1) Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 2) Department of Cell and Developmental Biology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Mutations in microRNAs (miRNAs) have been discovered to lead to deafness in both humans and mice. miRNAs play a critical role in development and regulation of sensory systems, including the inner ear, which is responsible for hearing and balance in mammals. Mutations are associated with human hearing impairment (Mencia et al. Nat Genet 2009). The study of hearing impaired mutants, however, has been hampered by the inability of inner ear RNA from human subjects, making the mouse an invaluable model for studying miRNAs in the inner ear. For example, removal of Dicer using the Cre-loxP recombination system, driven by a Pou4f3 hair cell-specific promoter, lead to the deletion of hear cell miRNAs and resulted in complete deafness by one month of age (Friedman et al. PNAS 2009). To further dissect the function of miRNAs in the mammalian inner ear, we performed RNA-Seq on RNA isolated from mouse inner ear sensory epithelia. qRT-PCR confirmed the expression of these miRNAs and in situ hybridization was used to identify their spatial expression in the mouse inner ear. We used bioinformatics software to predict the targets of these miRNAs and verified them by in vitro expression and luciferase assays. A total of 7,732,589 and 6,452,794 small RNA reads were found in the cochlear and vestibular samples, respectively. These included miRNAs, snoRNAs, transfer RNAs and ribosomal RNAs. Reads were aligned to the mature Mus musculus miRNA database (http://www.mirbase.org). miRDeep2 was used for novel microRNA prediction. We chose miRNAs for further study with a seed region that was not previously described between conserved and novel sequences of the non-coding region of the inner ear. One example is a miRNA gene that is included in the intron of the gene Tectb, which is expressed in the mouse inner ear and is associated with deafness. The gene produces two mature miRNAs, 5p and 3p, and both are upregulated in human alcoholics relative to controls.

Posters: Genome Structure, Variation and Function

Integrated microRNA and mRNA signature associated with the transition from the locally confined to the metastasized renal cell carcinoma. J. Biluadi, Z. Wotschokosz1, K. Jung2, H. Meyer1, 1) Ingenuity Systems, Redwood City, CA; 2) Department of Urology, Charité - Universitätsmedizin Berlin, Berlin, Germany.

MicroRNAs (miRNAs) are endogenous small non-coding RNAs that regulate gene expression by interfering translation or stability of target transcripts. One miRNA can interact with several hundred mRNAs, while one mRNA can be regulated by several miRNAs. This interplay between miRNA and mRNA can be regulated by several miRNAs. This interplay between miRNA and mRNA can be regulated by several miRNAs. Integrated miRNA-mRNA interaction analysis can provide valuable insights into the mechanisms underlying cancer progression. Here, we investigated miRNA-mRNA interactions in clear cell renal cell carcinoma (ccRCC) using novel experimental and bioinformatic approaches. We used Ingenuity pathway analysis microRNA Target Filter, which enables prioritization of experimentally validated and predicted mRNA targets. By applying an expression pairing tool, the analysis was focused on targets exhibiting altered expression in our analysis, finding miRNAs and their target genes with opposite or same expression. The resulting identified interactions were revalidated by RT-qPCR in another cohort of RCC patients. The predicted miRNA-mRNA interactions were also tested by functional analyses using miRNA knock-down and over expression experiments in renal cancer cell lines. Among the miRNAs differentially expressed, we have identified 3 miRNAs (miR-146a, miR-128a and miR-17-5p) that were upregulated in primary tumors from patients without metastasis and down regulated in primary tumors from patients with metastasis. We have further identified the mRNA targets which are targeted by these three 3 miRNAs, and have been previously experimentally demonstrated in cancer setting in humans. Specifically we showed that BRCA1, MCM10, CDKN3, UHRF1, IL8 were downregulated and targeted by miR-146a-5p. The relation between these miRNA targets and miRNAs is well conserved in cell culture experiments. In conclusion, we have identified novel target genes of miRNAs which are involved in the transition from primary RCC without metastases into tumors generating distant metastasis.
MicroRNAs (miRNA) are evolutionarily conserved regulators of protein expression, and are responsible for controlling transcription of up to 60% of expressed genes. For this reason, miRNA profiling strategies are increasingly being applied in clinical trials as biomarkers for complex diseases such as cancer and diabetes. Despite the important role of miRNA in disease, normal miRNA functional variation among different ethnic groups remains an open question, specifically in under-sampled populations such as ethnically and geographically diverse Africans. Here, we examine worldwide variation in miRNA genes by analyzing whole genome sequencing (60x coverage) in a panel of 69 unrelated individuals from 14 different ethnic groups. We include diverse African groups, not yet included in HapMap or 1000 genomes datasets, in an effort to identify novel human variation within miRNAs, as well as population-differentiated miRNA variants. We identified 198 novel variants within miRNAs, not present in dbSNP, with 29 novel variants located in the highly conserved seed sequence of mature miRNA. Additionally, between miRNAs, not present in dbSNP, with 29 novel variants located in the highly conserved seed sequence of mature miRNA. Many studies have focused on the relationship between miRNA and mRNA expression; thus, we sought to expand the regulatory understanding of the role of miRNA by evaluating its relationship to protein expression levels. We quantified 220 miRNAs and 441 protein isoforms across 68 LCLs derived from Yoruba cell lines from Ibadan, Nigeria. Two miRNAs, mir768-3p and mi29c, correlated with STAT3 protein expression with a Bonferroni-corrected p < 0.05. These relationships were not predicted bioinformatically or at the mRNA level. We then evaluated all miRNAs for the number of protein levels correlated with each miRNA. Using permutation analysis, each miRNA was found to be associated with more protein levels than would be expected by chance. This enrichment was robust independent of the significance level of the protein-miRNA relationship (p < 0.05, p < 0.01, p < 0.005). We identified two master regulatory miRNAs whose expression levels were associated with greater than ten other proteins at p < 0.001: mir125a-5p and mir768-3p. Expression of mir768-3p was significantly associated (p < 0.01) with response to five different chemotherapeutic agents (cisplatin, carboplatin, cytarabine, etoposide, and daunorubicin). Mir125a-5p has been widely implicated in the literature for its involvement in many complex processes, including carcinogenesis, inflammation, and metabolism. Our work annotates protein levels that may be regulated by microRNA. Further study of the role of microRNAs in protein expression will lead to a better understanding of dynamic cellular processes and pharmacologic phenotypes.

**Study of microRNAs regulated by hypoxia in cells latently infected by Kaposi’s sarcoma-associated virus.**

C. Viollet-Djelassi, C. Davis, C. Camps, M. Recekko, F. Pezzella, R. Yarchicoan, I. Ragoussi. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) HIV and AIDS Malignancy Branch, Center for Cancer Research, National Cancer Institute, National Institute of Health, Bethesda, MD; 3) BSRC Al. Fleming, Vari, Athens, Greece.

Kaposi’s sarcoma-associated herpesvirus (KSHV) infection, especially when combined with immunodeficiency, can lead to cancers such as Kaposi’s sarcoma (KS), primary effusion lymphoma or multicentric Castleman’s disease. KSHV is a highly vascular tumor, and KSHV genes have been shown to be induced by hypoxia and in turn to activate hypoxia-inducible factors, indicating a role of hypoxia in the KSHV life cycle. Additionally, virus-encoded microRNAs (miRNAs) have been increasingly recognised as contributors to viral cancer pathogenesis. Investigating their role in KSHV-related diseases is therefore of interest. We hypothesise that hypoxia and/or KSHV infection will alter cellular and viral miRNA levels that play a role in cancer progression and viral pathogenesis. To address this, we compared miRNA expression profiles of KSHV infected cells and non-infected cells under hypoxia and normoxia using Illumina small RNA and total RNA deep sequencing. In infected endothelial cells, we found that 112 mature human miRNAs and 4 viral miRNAs are differentially expressed in hypoxia compared to normoxia. In particular, has-mir-663b was the most down-regulated miRNA in hypoxia (p <= 0.05), while has-mir-210 was one of the most up-regulated (7.257-fold) and its increases in 1% O2 were further validated by qRT-PCR in different KSHV-positive cell lines (B-cells BCBL1 and endothelial cells SLKK). Interestingly, KSHV-mir-K12-3-3p, a known promoter of KSHV latency, was also significantly down-regulated in hypoxia compared to normoxia in endothelial SLKK cells. The effect of the hypoxia-related deletion of miRNAs on the expression patterns of both host and virus miRNAs is being elucidated through integrated analysis in order to identify targets and respective proteins being investigated in relation to KSHV-associated diseases and the hypoxia pathway. The outcomes of the present study will aid our understanding of how KSHV uses the host RNA silencing machinery to its advantage and provides clues as to how KSHV interacts with the human genome. This research was supported by the Wellcome Trust and the Intramural Research Program of the NIH, NCI.

**Functional Assessment of snoRNAs derived microRNAs in Prader-Willi Syndrome**

V. Williamson, M. Mamdani, G. McMichael, V. Vladimirov, Psychiatry, Virginia InstPsychiatric & Behavioral Genetics, Richmond, VA.

Prader-Willi Syndrome (PWS) is a neurodevelopmental disorder, characterised by hyperphagia, obesity and self-injurious behaviors such as skin-picking, temper tantrums and impulsivity. Approximately, 70% of subjects with PWS carry a deletion on chromosome 15 (q11.1-11.3) containing a small nucleolar RNA (snoRNA) cluster, HBII-85. Recently, snoRNA were shown to be induced by hypoxia and in turn to activate hypoxia-responsive genes. HBII-85 "snomiRs" targeting neuronal or brain development may be contributors to viral cancer pathogenesis. Investigating their role in KSHV-associated diseases is therefore of interest. We hypothesised that HBII-85 affects Prader-Willi syndrome development through HBII-85 derived miRNA. Many studies have focused on the relationship between miRNA and mRNA expression; thus, we sought to expand the regulatory understanding of the role of miRNA by evaluating its relationship to protein expression levels. We quantified 220 miRNAs and 441 protein isoforms across 68 LCLs derived from Yoruba cell lines from Ibadan, Nigeria. Two miRNAs, mir768-3p and mi29c, correlated with STAT3 protein expression with a Bonferroni-corrected p < 0.05. These relationships were not predicted bioinformatically or at the mRNA level. We then evaluated all miRNAs for the number of protein levels correlated with each miRNA. Using permutation analysis, each miRNA was found to be associated with more protein levels than would be expected by chance. This enrichment was robust independent of the significance level of the protein-miRNA relationship (p < 0.05, p < 0.01, p < 0.005). We identified two master regulatory miRNAs whose expression levels were associated with greater than ten other proteins at p < 0.001: mir125a-5p and mir768-3p. Expression of mir768-3p was significantly associated (p < 0.01) with response to five different chemotherapeutic agents (cisplatin, carboplatin, cytarabine, etoposide, and daunorubicin). Mir125a-5p has been widely implicated in the literature for its involvement in many complex processes, including carcinogenesis, inflammation, and metabolism. Our work annotates protein levels that may be regulated by microRNA. Further study of the role of microRNAs in protein expression will lead to a better understanding of dynamic cellular processes and pharmacologic phenotypes.

The olfactory neuroepithelium (ONE) represents a site where neural stem cells (NSC) can be obtained, as well as being very accessible. These cells can be cultured by neurosphere assay, forming spherical clusters of multipotent and progenitor cells in suspension with the ability to differentiate. These features make it a model for studying neural molecular and cellular processes. The ability of these cells to maintain an undifferentiated state and their capacity to differentiate requires that the cell can modify its expression, therefore depending on epigenetic mechanisms modulating molecules that promote the multipotentiality and that suppress cell differentiation. One such mechanism depends on microRNAs, small molecules of non-coding RNAs with roles in the regulation of expression. Its functions are both, neural lineage determination and in adult neural tissue functions. Obtaining human NSCs is a powerful tool to investigate neural processes. With this in mind, we cultured ONE cells from nasal swabs from 4 healthy volunteers. Neural marker expression was determined by by immunofluorescence. Subsequently a miRNA expression assay was performed with plates TLDA A y B of applied V 2.0. Cq values equal or greater than 36 were discarded. We subsequently a miRNA expression assay was performed with plates TLDA A y B of applied V 2.0. Cq values equal or greater than 36 were discarded. We assessed the expression level according to the DCT method, normalizing with the global mean and forming groups according to the mean plus / minus 1 or 2 standard deviations. Finally, we performed an in silico analysis looking for possible nervous system pathways in which expression of microRNAs are involved. We analyzed a total of 667 microRNAs. 246 microRNAs were expressed in at least 3 of the 4 samples. According to the expression level, 56 miRNAs have very high expression (90%), 23 high (9%), 66 medium (27%), 34 low (14%) and 67 very low (28%). Some of the microRNAs with higher expression are miR-222, miR-200c, miR-191, miR-30a *, miR-30e *, miR-484, miR-146a, miR-378, miR-24 miR-574-5p. Some miRNAs are already described with roles in neurogenesis as miR-184, miR-132 family members’ miR-200c, and miR-30, among others. As for the expressed miRNAs possible role, predictions point to a possible involvement in neurodevelopmental pathways as well as in neural pathologies.

617W LncRNAs regulation in insulin resistance of the adipose tissue. M. Pradas-Jun1,2, X. Bollf1, F. Hanzu1,2, C. Fillat4, R. Gomez1,2, E. Fernandez-Rebollo1,2, 1) Diabetes and Obesity Research Laboratory, IDIBAPS, Barcelona, Spain; 2) Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Diseases (CIBERDEM), Barcelona, Spain; 3) Endocrinology Unit, Hospital Clinic de Barcelona, Barcelona, Spain; 4) Gene Therapy and Cancer, IDIBAPS, Barcelona, Spain.

The non-coding RNAs (ncRNAs) have been largely underestimated but lately are being recognized as essential regulators of translational regulation and other processes. The ncRNAs are aberrantly expressed in a variety of human diseases; and thus demonstrating potential roles in cellular development and metabolism. NcRNAs are classified into two major classes based on transcript size: small and long ncRNAs (lncRNAs). LncRNAs are mRNA-like transcripts longer than 200 nucleotides, lacking significant open reading frames generally are transcribed by RNA polymerase II and are emerging as key regulators playing a major biological role in epigenetics, alternative splicing and even as regulators of mRNA decay. On the other hand, type 2 diabetes mellitus (T2DM) is the most common metabolic disorder in the world and obesity, meaning visceral adiposity, is the core problem. T2DM is characterized by hyperglycemia and impaired insulin action and/or secretion. The main role of obesity in T2DM is due to the visceral adipose tissue (VAT) insulin resistance. Based on this information we hypothesize that IncRNAs are key regulators in the insulin resistance of the VAT in T2DM. Aim: Characterize the IncRNAs and their ‘co-regulated’ mRNAs involved in insulin resistance of VAT. Methods: VAT from 8 obese and 8 obese with T2DM patients was used for RNA isolation to perform human IncRNA microarray V2.0 (Arraystar), which is designed for the global profiling of human IncRNAs (33,945) and protein-coding transcripts (30,215 mRNAs). To identify the highly correlated mRNA and IncRNA pairs, coding-non-coding gene expression network analysis (CNC) was performed. Results: First analysis identified 1,859 upregulated and 286 downregulated IncRNAs, and 1,136 upregulated and 238 downregulated mRNAs in T2DM obese patients compared to the obese controls (Fold Change ≥ 2.0, P-value ≤ 0.05). Specificity was increased by reducing the P-value (≤ 0.01) and obtaining 215 CNCs, and taking into account the IncRNA genomic localization in respect to mRNA. Using this data, we characterized the piRNA cluster distribution across the human genome and measured the extent of piRNA expression variation in human testis samples and generated over 50 million putative piRNA reads. From these data, we performed an in silico analysis looking for possible nervous system pathways in which expression of microRNAs are involved. We analyzed a total of 667 microRNAs. 246 microRNAs were expressed in at least 3 of the 4 samples. According to the expression level, 56 miRNAs have very high expression (90%), 23 high (9%), 66 medium (27%), 34 low (14%) and 67 very low (28%). Some of the microRNAs with higher expression are miR-222, miR-200c, miR-191, miR-30a *, miR-30e *, miR-484, miR-146a, miR-378, miR-24 miR-574-5p. Some miRNAs are already described with roles in neurogenesis as miR-184, miR-132 family members’ miR-200c, and miR-30, among others. As for the expressed miRNAs possible role, predictions point to a possible involvement in neurodevelopmental pathways as well as in neural pathologies.

618T Characterization of piRNA genomic distribution and expression variation in human individuals. J. Xing1,2, H. Ha1,2, J. Song1,2, S. Wang1,2, K.C. Chen1,2, 1) Dept of Genetics; 2) Human Genetic Institute of New Jersey; 3) BioMaPS Institute for Quantitative Biology; Rutgers, The State University of New Jersey, Piscataway, NJ.

Piwi-interacting RNAs (piRNAs) are a class of recently discovered small non-coding RNAs whose best known function is to repress mobile element activity in animal germlines. To date, virtually all piRNA studies have been conducted in model organisms and little is known about piRNA diversity, target specificity and the mechanism of mobile element regulation in humans. In this study, we performed high-throughput piRNA sequencing in three human tissues samples and generated over 50 million putative piRNA reads. Using this data, we characterized the piRNA cluster distribution across the human genome and measured the extent of piRNA expression variation among the three samples. Overall we identified ~10,000 piRNA clusters in the human genome. The piRNA clusters range from 1kb to 276kb in size and occupy ~3% of the genome. PiRNA clusters within genes are enriched in the 3' UTR region, consistent with previous findings in mouse. To examine the role of piRNAs in mobile element regulation, we determined the piRNA mapping density in the consensus sequence of Alu and L1 elements, the most abundant mobile elements in the human genome. We found that piRNAs preferentially mapped to specific motifs in the consensus of these mobile elements, and some piRNA mapping peaks show sequence/position patterns consistent with the ping-pong mechanism. In addition, we showed that human piRNA clusters show smaller variation in their expression level among human individuals than in Drosophila. Our study provides a comprehensive characterization of piRNA diversity and their interaction with genes and mobile elements in humans.
619F
MARK4 is a Ser-Thr kinase that phosphorylates MAPs taking part in the regulation of microtubule dynamics involved in cell cycle control. The MARK4 gene encodes two alternatively spliced isoforms: the canonical MARK4L, featuring 18 exons, and the alternative MARK4L4, derived from skipping of exon 16. In glioma we pointed out an imbalance between the MARK4L isoforms with decreased MARK4L4 expression associated with overexpression of MARK4L. A high L/S ratio also characterizes human glioblastoma-derived stem cells and mouse neural stem cells and appears proportional to cellular de-differentiation and tumor grade. Since the deregulation of MARK4 expression in glioma is not due to mutations or copy number loss/gain, we hinted that alterations in alternative splicing (AS) may be at the origin of the observed MARK4 isoforms imbalance. It’s well known that specific splice variants are commonly enriched in cancers as a consequence of splicing factors up-regulation. In glioma, in particular, overexpression of PTB, a key component in regulating neural stem cell proliferation and differentiation, drives an oncogenic splicing switch favoring isoforms, like MARK4L4, derived from exon skipping. Bioinformatic analysis of the MARK4 sequence by SFMap software revealed three putative PTB binding sites in both introns flanking exon 16. A functional role of these sites is suggested by the high conservation in mouse, as confirmed by CLUSTAL W alignment, and by the surrounding polypyrimidine rich context, required for PTB activity. Western blot analysis showed a significant overexpression of PTB in our astrocytoma and glioblastoma cell lines. In theory, the overexpression of PTB and the site with the highest PTB: L expression. Deletion of the last 80 nt of MARK4 IVS15 foster exon 16 inclusion in a splicing minigene system, revealing the presence of a functional intronic splicing silencer (ISS) in this region. However, mutagenesis of the predicted PTB binding site contained in the deletion does not affect expression. Deletion of the last 80 nt of MARK4 IVS15 foster exon 16 and by the surrounding polypyrimidine rich context, required for PTB activity. Deletion of the last 80 nt of MARK4 IVS15 foster exon 16 and by the surrounding polypyrimidine rich context, required for PTB activity. Deletion of the last 80 nt of MARK4 IVS15 foster exon 16 and by the surrounding polypyrimidine rich context, required for PTB activity. Deletion of the last 80 nt of MARK4 IVS15 foster exon 16 and by the surrounding polypyrimidine rich context, required for PTB activity.

621T
Role of RNA Editing in ER Stress Response. A. Richards1, I. Wang2, V. Cheung3. 1) Cell and Molecular Biology Graduate Program, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA; 3) Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA; 4) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA 19104, USA.
Adenose Deaminase Acting on RNA (ADAR) deaminates adenosine to inosine in double-stranded RNA. There are thousands of adenosine to inosine editing events in human cells. While some editing sites, such as those in ion channels, including the AMPA receptor, have been well studied, the role of RNA editing in a more general cellular context is less well understood. In this project, we aim to understand how RNA editing of genes involved in the endoplasmic reticulum (ER) and the Golgi affects their functions, particularly in ER stress response. We have identified ~1,000 human genes edited by ADAR, as confirmed by a decrease in editing levels following ADAR knock-down. Among them, 108 are genes localized to the endoplasmic reticulum and the Golgi body. To study these editing events, we sequenced the mRNA and corresponding DNA from B-cells of 10 individuals. We found about 710 editing sites within these 108 ADAR target genes; on average, there are 6 editing sites and a range of one to 54 sites per gene. The editing levels vary across individuals. For example, the editing level of a site in the 3’ UTR of EIF2AK2 varies by greater than 4-fold across individuals (range 13%-57%). In addition, the editing levels differ at different sites. To study if editing levels change following the cellular stress, we treated B-cells of 10 individuals with tunicamycin to induce ER stress. We then sequenced the DNA and mRNA of these cells before and at 2 and 9 hours following ER stress. About 275 of the 710 editing sites increased in editing level >1.5-fold following ER stress. Editing level changes following ER stress also showed individual variability. For example, the editing level of a site in the 3’ UTR of VHL increases 1.8- to 7-fold across individuals following ER stress. Over 60% of ADAR editing sites in the ER and Golgi genes are located in the 3’ UTR. Therefore, we aimed to determine if RNA editing influences gene expression to affect ER stress response. We used luciferase reporter constructs with the edited sequence of 3’ UTRs and found that regions of variable editing in the 3’ UTR can modulate reporter activity. In this presentation, we will illustrate the features of RNA editing in ER and Golgi genes. We will, further, describe how RNA editing levels respond to ER stress and how editing, in turn, may impact ER stress response.

622F
Musashi1 (Msi1) is an evolutionarily conserved RNA-binding protein that has been implicated in a variety of cellular processes such as stem cell maintenance, nervous system development, and tumorigenesis. Msi1 is highly expressed in many cancers including glioblastoma, and is emerging as a potential therapeutic target in both regenerative medicine and cancer. Our goal is to better understand the regulatory role of Msi1 by identifying the RNAs targeted by Msi1 at the level of translation. Our approach is to use high-throughput RNA sequencing (RNA-Seq) to compare total mRNA to the Ribosome Protected mRNA Fragments (RPF) in order to provide quantitative analysis of translationally active mRNAs versus repressed pools of mRNAs present at a particular time or condition. Consequently, we developed a simplified method for preparing RNA-Seq libraries from Ribosome Protected fragments. A small hairpin (sh) RNA was used to knock down Msi1 activity in the glioblastoma cell line U251. RNA-Seq libraries were prepared from the ribosome protected fragments of control and shMsi1-treated cells and sequenced on the Illumina platform. Our results show RNA-Seq of ribosome protected fragments offers a viable approach to understand the translational regulation role of Msi1.
623W
Integrator Complex Subunit 8 mutation associated with cortical and cerebellar malformations results in disruption of the Integrator complex and sp无需完善，...[等]。我们还发现了与这些变异相关的PHEX和FGF4基因的异常表达，这可能与表达异常相关的神经发育障碍有关。这些结果表明，Integrator Complex Subunit 8 (INTS8) 可能与神经发育障碍有关，并且可能在人类的发育障碍中发挥作用。

625F
Mutation profiling of exonic-enhancers using massively parallel reporter assays, R.Y. Bimbaum, R.P. Patwardhan, M.J. Kim, G. Finlay, D. Zhao, R. Beil, R.P. Smith, A.A. Ku, J. Shendure, N. Ahituv. 1) Department of Bioengineering and Therapeutic Sciences, Institute for Human Genetics, University of California San Francisco, San Francisco, California, USA; 2) Department of Genome Sciences, University of Washington, Seattle, Washington, USA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94143, USA.

Protein coding exons that also function as enhancers (eExons) have been shown to regulate transcription of the genes they reside in or nearby genes. Mutations in these eExons could lead to multiple phenotypes due to alterations in protein function and/or transcriptional regulation. However, the functional consequences of these mutations are not well known. Here, using ChiP-seq with enhancer labels (p300, H3K27ac, H3K4me1) on human hepatocytes and mouse liver, we show that ~6% of all ChiP-seq peaks overlap coding exons (excluding 1st exon and promoter regions) and demonstrate that 8 of 15 tested sequences function as enhancers in mouse liver and HepG2 cells. Using massively parallel reporter assays, we further dissect in vivo the enhancer activity of these eExons across cell types leading to differences in deleterious mutation profiles. Combined, these results demonstrate that eExons mutations can disrupt both the protein structure and enhancer activity with differential effect across cell types and can cause multiple phenotypes.

626W

Down syndrome (DS) results from trisomy of chromosome 21 (HSA21). Some DS phenotypes may be directly or indirectly related to the increased expression of specific HSA21 genes, in particular transcription factors. The HSA21 Single-minded 2 (SIM2) transcription factor has key neurological functions and appears therefore as a good candidate for some DS features, in particular mental retardation. In order to identify DNA binding sites and downstream targets of SIM2, we sequenced SIM2-immunoprecipitated DNA from a mouse embryonic stem cell (mESC) line overexpressing a Flag-tagged mouse Sim2 under the control of a Tet-off system. Non-expressing mESCs were used as controls. Reads uniquely mapped with BWA were then mapped to the mouse genome (Ensembl database). Interestingly, these areas, from which neuronal precursors migrate to the cortex, are compatible with the regions affected by PNH and cerebellar hypoplasia in the patients. We propose that dysfunction of the Integrator Complex, possibly through snRNA misexpression and/or splicing defects, leads to severely disrupted brain development in humans.
627T  
Splicing QTL analysis from primary immune cells identifies regulatory effects putatively associated with autism and Alzheimer’s disease. J.M. Replinger1,2,3, T. Raj1,2,3,4, E. L. Rothamel1, C. Benoit1,2,3,4, B. E. Stranger1,2,3,4, P. L. De Jager1,2,3,4, Immunological Variation Consortium. 1) Department of Neurology, Brigham & Women’s Hospital, Boston, MA; 2) Program in Medical & Population Genetics, The Broad Institute, Cambridge, MA; 3) Harvard Medical School, Boston, MA; 4) Division of Genetics, Department of Medicine, Brigham & Women’s Hospital, Boston, MA; 5) Department of Microbiology and Immunobiology, Harvard Medical School; 6) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 7) Institute of Genomics and Systems Biology, University of Chicago.

Alternative splicing acts as an abundant source of transcriptional and phenotypic diversity in humans. Recently, advances in sequencing and array technologies have facilitated high-throughput quantification of mRNA expression at the exon level, and previous studies highlight significant tissue- and population-specificity of alternative splicing. Several splicing quantitative trait loci (sQTL) studies have identified putative functional links between variants associated with complex traits and alternative splicing. However, many of these studies have identified sQTLs in immortalized cell lines, and we have little knowledge about how these discoveries will translate to primary cell-types, which may be most relevant for unraveling disease phenotypes. Here we performed an sQTL analysis in two primary immune cell-types, CD4+ and CD14+ cells, from individuals of African ancestry, with alternative splicing of $CD33$ rs3865444, previously associated with Alzheimer’s disease susceptibility, correlates with alternative splicing of $CD33$. Further, we measured the sexual dimorphic expression of X-linked, Y-linked, and autosomal genes in a wide range of human tissues based on more than 500 RNA-seq datasets. A standard student-t test with step-down Benjamini-Hochberg correction at 5% FDR was applied throughout to assess genes with at least 1.2 fold expression difference between the sexes. We selected tissues with at least 15 individuals per tissue and similar number of female and male individuals. Genes located within the largest human pseudoautosomal region (PAR1) had significantly higher expression in male tissues, possibly due to spreading of X inactivation in females. In contrast, escape genes outside the PAR had significantly higher, but rarely doubled, expression in females compared to males, in a tissue-specific manner. This female bias was most pronounced for genes known to escape X inactivation but was also seen for other X-linked genes perhaps due to partial escape in specific tissues. Such female-specific bias was particularly pronounced in the gyrus and the cortical regions of the brain, in the skin, in CD4+ positive blood cells, and in reproductive organs. These studies led us to identify the first comprehensive sex biased human transcriptome in brain, non-brain, immune cells, and sex organs, which consisted of a female biased gene set containing 279 X-linked and 7229 autosomal genes (21% of the genome) and a male biased gene set containing 95 X-linked and 2015 autosomal genes (9% of the genome).
630T
FREM1 Regulates Genes Important for HIV-1 Replication and Cell Migration.
M. Luo1,2, J. Sainsbury1, P. Lacap1, F. Plummer1,2.
1) HIV and Human Genetics, Public Health Agency of Canada, Winnipeg, Manitoba, Canada; 2) Department of Medical Microbiology, University of Manitoba, Winnipeg, Canada.

Introduction: FREM1 has been identified as a novel candidate gene in resistance to HIV-1 infection in the Pumwani sex worker cohort established in Nairobi, Kenya. Several molecular features of FREM1 suggest its potential role in HIV-1 vaginal transmission. It is an essential component for epidermal integrity and in the right path of HIV-1 infection. A splice variant of FREM1 is a co-receptor for IL-1R1 and toll-like receptor involved in enhancing NF-kB activation.

However, the precise role of FREM1 in HIV-1 infection and how its variants influence resistance and susceptibility to HIV-1 infection need to be investigated. In this study we studies the role of FREM1 in regulating genes important in HIV-1 replication and cell migration, the two important factors in vaginal HIV-1 transmission.

Method: We knocked down FREM1 expression in 293F cells and over expressed FREM1 in the HeLa cells derived from cervical tissue and examined the effect on the expression of genes in signal transduction pathways important for HIV-1 transmission by real time PCR array (SA Bioscience).

Results: The results showed that knocking down or over expression of FREM1 influences the expression of many important genes involved in NF-κB and inflammatory response, apoptosis, epithelial adhesion and cell migration. As immune response is of paramount importance to HIV-1 transmission and replication, regulation of immune system genes by FREM1 may be a major factor in mediating resistance to mucosal acquisition of HIV-1.

Conclusion: These suggest that FREM1 may play an immunomodulatory role in cellular activation, which is critical for HIV-1 replication. FREM1 may act as a target and mediator for the development of mucosal barrier resistance to HIV-1 infection at vaginal mucosa as suggested by its potential role in epithelial integrity and cell migration through its different functional domains.

631F
In vivo UAS<sup>Gal</sup> gene regulation analysis using the a novel approach.
H. Guillen Ahlers1, A. Ludwig-Kubinski1, K. Lazarova1, A.M. Greene1, J. Kennedy-Darling2, M. Levenstein2, C. Anderson1, J. Barfknecht1, R. Knoerer1, M. Scal1, F. Yuan2, R. Cole1, M. Shortreed1, L. Cirillo1, L. Smith2, M. Olivier1, 1) WI CEGS, Medical College of Wisconsin, Milwaukee, WI; 2) WI CEGS, University of Madison, Madison, WI.

The Wisconsin Center for Excellence in Genomics Science is developing a novel technology to identify the proteins that are associated with DNA at any desired region of the genome. Sequence-specific hybridization to formaldehyde-crosslinked chromatin is used to capture the target region along with its interacting proteins. The novel approach builds on the previous GENECAPP technology developed by the present group and does not require restriction enzyme digestion. The new approach utilizes an optimized capture oligonucleotide mixture that greatly increases the efficiency of the process while reducing costs. This technology was used to study the Gal upstream activator sequence (UAS<sup>Gal</sup>) in S. cerevisiae. Cells were grown using either glucose or galactose as the carbon source, so that the difference in gene modulation at the UAS<sup>Gal</sup> region could be observed. Bound proteins were identified by tandem mass spectrometry using an Orbitrap Velos instrument.

The development of this technology has shown steady improvements and highlights the great potential offered for an unbiased analysis of DNA-protein interactions. Funded by the Wisconsin CEGS through NIH/NHGRI grant 1P50HG004952.

632W
Functional impact of polymorphic inversions on gene expression in humans.
M. Oliva1, S. Villatoro1, L. Pantano1, C. Aguado1, D. Vicente-Salvador1, D. Izquierdo1, M. Puig1, T. Marques-Bonet2,4, JI. Lucas Lledó1, R. Castelo3, M. Cáceres1,2,4.
1) Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain; 2) Institut de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra), PRBB, Barcelona 08003, Spain; 3) Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain; 4) Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

Despite the significant advances made over the last few years, our understanding of the prevalence and functional impact of inversions in the human genome is scarce. Moreover, little is known about associations that influence gene expression changes in humans. The only exceptions are a few studied examples, such as the association of the 17q21.31 inversion with decreased MAPT expression or the 8p23.1 inversion with decreased PPP1R3B expression. Here, we analyzed the overlap of a set of polymorphic inversions in the human genome with genes and the correlation between inversion genotypes and gene expression profiles in a genome-wide fashion, looking for both cis and trans effects. First, we recorded the gene content of 48 human inversions (identified by our group) and generated 1000 permuted sets to build an empirical null distribution of gene counts for each inversion. We confirmed that inversions break fewer genes than expected by chance (p<0.05). Next, we looked for associations of 37 validated inversions with gene expression changes by analyzing microarray expression datasets of lymphoblastoid cell lines of 90 Yoruba, 90 European, 45 Chinese and 45 Japanese HapMap individuals. For each gene expression profile, we searched for inversion eQTLs using linear models in which we adjusted for confounding factors using surrogate variable analysis (SVA) and tried to quantify the effect of each inversion. Possible influencing variables (gender, population) were also taken into account in the model. We further analyzed the two well-characterized inversions (17q21.31, 8p23.1) to evaluate the reliability of our methodology in detecting known effects of inversions on the expression of particular genes. Our results show that 30% (11/37) of studied inversions associate to the expression of at least one gene in cis (FDR<10%) and a subset of them also seem to affect expression of genes in the genome. Additional analyses of expression datasets derived from alternative splicing, of RNA-Seq were performed to assess the robustness of our findings. Insight gained in this study could contribute to a better understanding of the role of polymorphic inversions in the regulation of gene expression that may be relevant for the human genome.
Targeted sequencing of promoter-associated tandem repeats identifies common functional effects on gene expression levels in the human genome. A.J. Sharp1, A. Guilmatre1, P. Garg1, G. Highnam2, D. Mittelman2,3. 1) Genetics & Genomics Sci, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA, USA; 3) Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA.

Tandem repeats (TRs) are stretches of DNA comprised of two or more contiguous copies of a motif arranged in a head-to-tail pattern, and comprise ~2% of the human genome. They are characterized by high mutation rates and account for >25% of indel variants, therefore representing an important source of genetic variation. However, due to technical difficulties in studying them, TRs remain poorly studied and have often been considered as mere 'junk DNA'. We hypothesized that many TR variations might operate as expression quantitative trait loci (eQTLs). To investigate this hypothesis we have conducted a cis-association analysis of variation in promoter-associated TRs with neighboring gene expression levels. Utilizing a custom solution-based capture approach (PMID: 23696428) we studied 120 CEU and YRI HapMap individuals, obtaining a median coverage of 47x informative reads for >4,000 TRs located within 1kb of RefSeq gene transcription start sites. TR genotypes were called using the RepeatSeq algorithm (PMID: 23090981). We then performed an eQTL association analyses using published RNAseq data, identifying 432 TRs in the CEU and/or YRI populations that show significant association (p<0.05) with the expression level of adjacent genes. This included many TR loci that have not previously been reported as polymorphic. TRs scored as significant eQTLs were enriched for overlaps with transcription factor binding sites and putative enhancers, providing a strong biological rationale for their effects. After phasing of SNP and TR genotypes using BEAGLE, we analyzed patterns of linkage disequilibrium between TRs and nearby SNPs, and observed that most TR variants are poorly tagged by SNP markers. Only 8% of TRs had R2>0.8 with any SNP within 250kb, and the majority had R2<0.3 with the best tagging SNP. Thus although many TR variants show clear evidence of functional effects, this indicates that the majority are not effectively assayed by SNP-based GWAS approaches, potentially explaining some of the 'missing heritability' of the genome. Our eQTL study represents the first systematic attempt to assign biological significance to TR variations in the human genome, and suggests that potentially there are many thousands of TR variations in the genome that exert functional effects via alterations of local gene expression or epigenetics. We conclude that specific studies that focus on genotyping TR variants are required to fully ascertain functional variation in the genome.

Targeted CD4+ Effector Memory T-Cell Gene Expression Profiling Identifies State-Specific cis-eQTLs Among Rheumatoid Arthritis and Celiac Disease Risk Variants. X. Hu1,2,3, H. Kim1, C. Baecher-Allan1,6, T. Raj1,6, P. Brennan1,2,4, P. De Jager1,6, M. Brenner1,2,4, S. Raychaudhuri1,2,4,5,7,8. 1) Division of Rheumatology, Immunology and Allergy, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA; 2) Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA; 3) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 4) Harvard Medical School, Boston, MA, USA; 5) Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA; 6) Department of Neurology, Brigham and Women’s Hospital, Boston, MA, USA; 7) Partners Center for Personalized Genetic Medicine, Boston, MA, USA; 8) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

Motivation: We recently demonstrated that common single-nucleotide polymorphism (SNP) variants associated with rheumatoid arthritis (RA) and celiac disease (CeD) both implicate genes specifically expressed in CD4+ effector memory T (TEM) cells. We hypothesize that some variants may be expression quantitative trait loci (eQTL) that contribute to the development of disease by regulating gene transcription in a cell-specific manner. Methods: We performed genome-wide SNP analysis in 174 healthy, non-Hispanic, Caucasian volunteers. In conjunction, we isolated highly purified CD4+ TEM cells (CD45RA-, CD45R0high, CD62Llow/-) from the peripheral blood of the same individuals, and assayed the expression of 215 genes within RA and CeD loci along with 15 control genes for calibration using Nanostring nCounterTM, before and after T cell receptor (TCR) stimulation by anti-CD3/anti-CD28 antibodies. We assessed the presence of cis-eQTL within 1Mb of the transcription start site changes of each gene. Results: The genes with the largest fold changes in expression following stimulation were GZMB (average fold change: 182.2; range: 6.5-573.4) and IL2RA (average fold change: 105.9; range: 8.7-711.0). The genes with the greatest fold change were also those most specifically expressed in CD4+ TEM cells (Spearman rho = 0.26, P = 5.4x10-4). Six of the 35 RA variants in high-density genotyped regions, and four of the 50 CeD variants were in tight linkage disequilibrium with cis-eQTLs (permutation-based P < 0.01, R2 > 0.7). One disease-associated eQTL (rs1980422/CD28, shared between RA and CeD) influenced CD28 expression specifically in cells before stimulation; rs12936049 and rs4840565 were associated with differential expression of GSDMB and BLK, respectively, both before and after stimulation. Remaining disease-associated eQTLs were associated with differential expression only after stimulation. None of the eQTL SNPs were significantly associated with the proportion of CD4+ TEM cells or with cell proliferation. Conclusion: Most RA- and CeD-associated eQTLs are specific to either resting or stimulated CD4+ TEM cell state. Genetic variants may contribute to the development of autoimmune diseases by regulating gene transcription in specific cell populations such as CD4+ TEM cells.
635W

Mapping the genetic architecture of gene regulation in whole blood in the KORA study. K. Schramm1,2, C. Marzi3, C. Schurmann4, M. Carstensen1, E. Reinmaa5,6, J. Cieger7, E. Mihhailov8,9, R. Mägi8, A. Peters6,7,8, K. Stauch6,12, M. Roden13, T. Ilg14, T. Meiting15,16, A. Metspalu6,7, C. Herder2, H. Graffel2, H. Prokisch1,2, J. K. Schramm1,2, Institute of Human Genetics, Helmholtz Zentrum Münchhen, German Research Center for Environmental Health, Neuherberg, Germany; 2) Institut für Humangenetik, Technische Universität München, München, Germany; 3) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 4) Ernst-Moritz-Arndt-Universität Greifswald, Interfaculty Institute for Genetics and Functional Genomics, Greifswald, Germany; 5) Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany; 6) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 7) Estonian Genome Center, University of Tartu, Tartu, Estonia; 8) Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 9) Estonian Biocenter, Tartu, Estonia; 10) Munich Health Alliance, München, Germany; 11) Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 12) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, München, Germany; 13) Department of Metabolic Diseases, University Hospital Düsseldorf, Heinrich-Heine University, Düsseldorf, Germany; 14) Medical School Hannover, Hannover Unified Biobank, Hannover, Germany.

Introduction: Analysis of whole genome expression quantitative trait loci (eQTLs) identified regulatory relationships at a genome-wide scale and thus for identifying regulatory pathways affecting disease susceptibility and other relevant traits. Methods: We performed eQTL analyses in 890 randomly selected fasting participants from the population-based KORA F4 study and replicated the results in independent non-fasting samples (EGUT, N=842). Linear regression models using additive effects with adjustments for principal components were applied. Genome-wide statistical significance was defined as 0.05/number of tests performed. Results: A total of 16 cis-eQTLs (defined as 50k-eQTLs) showed genome-wide significance. Of those, 87% and 80% of the cis- and trans-eQTLs, respectively, were confirmed in the replication samples. (The replication in a second cohort is ongoing and looks similar.) Network analysis for validation using Ingenuity Pathway analysis software identified signeificant eQTLs using Ingenuity Pathway analysis software identified an enrichment of pathways involved in the development and the activity of the immune system and a central role of the HLA-system. Furthermore, for the set of significant cis-eQTLs we observed (1) an overlap of 19% of genes detected in genome-wide analysis studies (GWAs) and recorded in the GWAS catalog so far (http://www.genome.gov/gwastudies, July, 18th, 2012), (2) major cross-tissue similarity (46–70%) with previously published cis-eQTLs found in monocytes, LCLs, lymphocytes, lung tissue, and liver tissue, (3) five chromosomal regions with simultaneous impact on multiple gene expression levels. Amongst the set of significant trans-eQTLs, a triangular relationship between an eQTL-SNP residing in a gene desert on chromosome 6q24, the gene expression probe of a known type 2 diabetes susceptibility gene (IGF2BP2), and adiponectin was identified. Conclusion: The present study identified numerous eQTLs in whole blood in a large Western European sample and provided evidence that these results offer a valuable resource for investigators studying the genetic architecture of regulatory pathways in whole blood. The high replication rate also in non-fasting subjects demonstrates the robustness of the regulatory effects in whole blood. Furthermore whole blood seems to be an informative tissue for an abundance of transcriptional regulatory relationships also in other tissues.

636T

Paired eQTL analysis of monocytes and differentiated macrophages. S. Makino1, V. Narangbal1, J. Knight1, B. Fairfax1,2, 1) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 2) Oxford Cancer Centre, Churchill Hospital, Oxford, OX3 1LJ.

Monocytes and macrophages form crucial cellular subsets of the innate immune system. Monocytes are short-lived circulating cells that upon migration into inflamed or damaged tissue differentiate into resident macrophages, promoting either resolution or chronicity of inflammation. Multiple eQTL analyses have determined that a proportion of eQTL exhibit high cellular specificity. The degree to which the activity of genetic determinants of gene expression is retained during cellular differentiation is unclear however. Here we aimed to investigate whether eQTL observed in CD14+ monocytes are maintained after their differentiation into macrophages. Primary CD14+ monocytes were isolated from peripheral blood mononuclear cells of 64 healthy Europeans. They were ex vivo differentiated into macrophages over 18 days with Tumour Necrosis Factor (TNF) and Granulocyte-Macrophage Colony Stimulating Factor (GMCSF). Differentiation into a macrophage phenotype was confirmed with quantitative PCR (qPCR) of the monocyte and macrophage markers, CD14, CD68 and CD163. Gene expression was subsequently analyzed with total RNA from monocytes and macrophages using Illumina HumanHT-12v4 BeadChips. Individuals were genotyped at 730,000 markers using Illumina OmniExpress v1 beadchips. eQTL analysis was performed using a linear model incorporating expression principal components as covariates. Here we report the results of this analysis. Major observations include the diminishment of the previously reported master regulatory region at 12q15 at the LYZ-YEATS4 locus upon macrophage differentiation, a result validated with parallel qPCR of LYZ. This study is aiming on identifying potentially interesting virus-host interactions responsible for CMV latency. The efficient replication and to maintain lifelong latency in immunocompetent hosts, CMVs have evolved numerous molecules mediating immune evasive properties, targeting both innate and adaptive immune responses. One of the striking immune evasive strategies is to interfere with JAK/STAT signal transduction, block IFN-stimulated gene (ISG) expression following viral gene expression, also during an initial ISG activation phase. CMVs encode more than 100 genes that are nonessential for growth in vitro and hence are likely to modulate the virus-host interaction in vivo, including immune modulatory genes. UL23 gene, encoding pUL23 protein, is one member of the human cytomegalovirus (HCMV) US22 gene family. The intracellular target of the pUL23 protein was investigated by using a yeast two-hybrid screening system with pUL23 as bait. Following the two-hybrid screen, a list of interesting interactors was generated and the interactions validated using Western blots using commercial antibodies recognizing the interactors identified. From these results, we are focusing on specific partners associated with STAT1 and IFN receptor. Glutathione S-transferase pull-down experiment revealed some interactors associated with HCMV pUL23 protein with a higher affinity than STAT-1. Therefore, it is suggested that pUL23 protein has the ability to interact strongly with host proteins and consequently to bring about the disruption of the complex formed from STAT-1 and the IFN receptor, probably resulting in suppression of the IFN signal transduction pathway.
638W
Resolving regulatory genetic variants in severe sepsis due to community-acquired pneumonia by mapping context-specific expression quantitative trait loci. E.E. Davenport1, J. Radhakrishnan1, P. Hamburger2, T. Mills3, P. Hutton1, C. Garrard2, C. Hinds2, J.C. Knight1. The GainS Investigators. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Adult Intensive Care Unit, John Radcliffe Hospital, Oxford, United Kingdom; 3) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, London, United Kingdom.

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

639T
A comparative transcriptome analysis identifies FGF23-regulated genes in HEK293 cells stably expressing KLOTHO. S. Diener1, T. Schwarzmayr1, A. Schmittfull1, T. Wieland1, B. Lorenz-Dempereux1, T.M. Strom1,2. 1) Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Bavaria, Germany; 2) Klinikum Rechts der Isar der Technischen Universität München, Institute of Human Genetics, Munich, Bavaria, Germany.

Phosphate homeostasis is regulated in a complex process that involves the interplay of different organs, tissues and systems. A key regulator of phosphate metabolism are necessary to identify possible therapeutic targets. For this purpose, we established an KLOTHO (HEK293-KL). To find differentially expressed FGF23-induced transcripts, we performed whole transcriptome analysis. We used the technology of RNA-Seq, which is a massively parallel sequencing approach to allow genome-wide analysis of gene expression profiles at a far higher resolution than available with microarray-based methods. Genome-wide transcriptional changes in HEK293-KL cells specifically caused by FGF23 were defined by comparing the transcriptome of FGF23-induced HEK293-KL cells with the transcriptome of not induced HEK293-KL cells. We tried to find novel genes that might belong to a network of factors involved in the regulation of phosphate homeostasis.

640F
Circulating miRNAs associated with High Altitude Sickness at the Qinghai-Tibetan Plateau. NE. Buroker1, X-H. Ning2, Z-N. Zhou3, K. Li4, W-J. Chen4, K-F. Wu5, W-Z. Zhu6, CR. Scotti7, SH. Chen1. 1) Pediatrics, 356320, University of Washington, Washington, SE; 2) Division of Cardiology, Seattle Children’s Hospital. Institute. Foundation, Seattle, WA; 3) Laboratory of Hypoxia Physiology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; 4) People’s Hospital of the Tibet Autonomous Region, Lhasa, China; 5) Center for Cardiovascular Biology and Regenerative Medicine, University of Washington, Seattle, WA.

Circulating miRNAs isolated from dried blood spots (DBS) were found to be associated with high altitude sickness (HAS) patients in Tibet. HAS arises from two different diseases which are acute (AMS) and chronic (CMS) mountain sickness. Circulating miRNAs differences were found between AMS Han Chinese patients and normal Han controls and between CMS Tibetan Chinese patients and normal Tibetan controls. HAS arises from hypoxia which affects some high altitude inhabitants or visitors and not others. The difference results from each individual’s genetic makeup where hypoxia related genes have been shown to be a major contributor to these sicknesses. Several fold changes increases (up regulation) were found in the hypoxia associated miRNAs let-7i-5p, miR-8-5p, miR-19a-3p, miR-23a-3p, miR-98-5p, miR-125a-5p, miR-181b-5p, mir-202-3p, miR-372, miR-381-3p, miR-519d, miR-520d-3p, and miR-656 for both HAS groups compared to their controls. Other miRNAs (miR-19a-3p, 302c-3p and 875-5p) were found to be up regulated in one HAS group and down regulated in the other HAS group indicating the genetic differences between the two sickness groups.

641W
Genetic and epigenetic regulation of human LincRNAs gene expression variation. K. Popadin, M. Gutierrez-Arcelus, E.T. Dermitzakis, E.S. Antonarakis. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland. Large intergenic non-coding RNAs (lincRNA) are still poorly functionally characterized in spite of the fact that they represent at least one fourth of the number of protein-coding genes in the human genome. To provide transcriptome-wide description of human lincRNAs, we have analyzed the natural variation of lincRNA expression levels by RNA-Seq (10–50 M reads) as well as the genetic (Illumina 2.5M Omni chip) and epigenetic (450K Illumina Infinium HD Methylation Assay) regulation of lincRNAs in the GenCord collection of three cell types (fibroblasts, lymphoblastoid cell lines and T-cells) from 195 unrelated European individuals. We have observed eight hallmarks of lincRNA functionality: (i) a negative correlation between the lincRNA gene size and their level of expression; (ii) high conservation score of ubiquitously expressed lincRNAs; (iii) genomic co-localization of expressed lincRNA genes with protein-coding genes involved in zinc-ion binding; (iv) a higher abundance of cis expression Quantitative Trait Loci (cis-eQTLs) in lincRNAs compared to protein-coding genes and the prevalent localization of these cis-eQTLs very close to Transcription Start Sites (TSS) of lincRNAs; (v) regulation of lincRNA expression by genetic variation independence of regulation of the neighboring protein-coding genes; (vi) an independent transcription when lincRNA and protein-coding genes found near each other; (vii) epigenetic regulatory patterns similar to protein-coding genes: presence of both positive and negative correlations between DNA methylation and gene expression, with negative correlations being closer to the TSS, and similar landscape of passive and active roles of DNA methylation in gene regulation; (viii) an enrichment of frequently expressed lincRNA genes with protein-coding genes involved in zinc-ion binding; (ix) a lower expression than protein-coding gene of almost all lincRNAs measured and (x) extensive landscape of lincRNA expression regulation.
642T Heritability of Gene Expression Levels in Genome-Wide Analyses. T. Huan1, C. Liu1, R. Joehanes1,2, X. Zhang1, M. Larson2, B. Chen1, C. Yao1, A. Johnson1, P. Munson2, P. Couchesne1, C. O’Donnell1, D. Levy1. 1 Division of Intramural Research, National Heart, Lung and Blood Institute; the NHLBI’s Framingham Heart Study; 2) Mathematical and Statistical Computing Laboratory, Division of Computational Bioscience, Center for Information Technology, NIH; 3) Department of Mathematics and Statistics, Boston University, Boston, Massachusetts.

 Genome-wide expression quantitative trait locus (eQTL) mapping studies reveal common genetic variants regulating gene expression. In addition to eQTLs, we systematically evaluated the heritability (h²) of the whole blood transcriptome in Framingham Heart Study (FHS) families, and explored the proportion of the heritability of gene expression explained by cis- and trans- components, as a means of identifying the role of eQTLs in promoting phenotype differences and disease susceptibility. Peripheral whole blood samples were collected and large-scale transcriptomic microarray measurements were performed on 5,626 FHS participants. The pedigree structure consisted of 704 extended pedigrees (of size ≥2) from two-generations of FHS participants. Heritability estimates of ~18,000 transcripts were obtained by variance-component methodology. Of all transcripts, about 40% displayed heritability h²>0 with p<0.05, and 10% displayed h²>0.2 (p < 1.8e-6). We then investigated the proportion of cis-trans-eQTLs in each of the heritability categories at h² = 0.2-0.29, 0.3-0.39, 0.4-0.49, 0.5-0.59 and ≥0.6. We discovered that transcripts with higher heritability estimates tended to have larger proportion of cis-eQTLs. In contrast, there was no apparent trend for the proportions of trans-eQTLs in the different heritability categories. In addition, we discovered that single cis-eQTLs explain 27-66% of variance in gene expression levels with h²>0.2. However, trans-eQTLs only account 1-6% of variance in transcripts with h²>0.2. Interestingly, we observed that the top cis-eQTL tended to explain more variance in the respective transcript as the heritability of the transcript increased, but the top trans-eQTL tended to explain more variance in the respective transcript when the heritability of the transcript decreased. By cross-linking the eQTLs with GWAS results of multiple metabolic traits, we discovered that a few trait-associated GWAS single nucleotide polymorphisms (SNPs) explained a large proportion of variance in the respective trait. We found that several differentially expressed genes were associated with the same GWAS SNPs. These gene signatures explained a larger proportion of variance in the respective traits than did the GWAS SNPs.

643F Cis and Trans Effects of Human Variants on Gene Expression. J. Bryois1,2, A. Bull1,2, D.M. Evans1,2, J.P. Kemp1,2, S.B. Montgomery1, D.F. Conrad2, K.M. Ho2, S. Ring2, M. Hurles2, P. Deloukas2, G.D Smith2,3, E.T Dermitzakis1,2,3. 1) Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Institute of Genetics and Genomics in Geneva (iGE3), Geneva, Switzerland; 3) Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland; 4) MRC Centre for Medical Genetics, University of Cambridge, Cambridge, UK; 5) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 6) Welcome Trust Sanger Institute, Hinxton, United Kingdom; 7) Departments of Pathology and Genetics, Stanford University, Stanford, California, United States of America.

Gene expression is a heritable cellular phenotype that defines the function of a cell and can lead to diseases in case of misregulation. In order to detect human variations affecting gene expression, we performed genome-wide association analysis of single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) with gene expression measured in 869 lymphoblastoid cell lines in cis and in trans. Using a threshold defined by permutations, we observed that 3925 genes (false discovery rate (FDR)=4%) are affected by expression quantitative trait locus (eQTL) in cis and 83 genes (FDR>20%) are affected in trans. We found that CNVs are more likely to be eQTLs than SNPs (odd ratio for cis-eQTLs=7.62 (pval=2.2e-16), odd ratio for trans-eQTLs=44.1 (pval=1.7e-7)). In addition we found that GWAS SNPs are enriched for cis and trans eQTLs (pval<0.01) and that trans-eQTLs are enriched for cis-eQTLs (pval<0.01). As a variant affecting both a gene in cis and in trans suggests that the cis gene is functionally linked to the trans gene expression, we looked specifically for trans effects of cis-eQTLs. We estimated the proportion of probes affected by each cis-eQTL from the pvalue distribution of their trans effects and discover that many of them have pleiotropic effects. Using a threshold defined by permutations, we found that 51 cis-eQTLs are associated to 151 genes in trans (FDR<1%) with the cis-eQTLs of the transcriptions factors BATF3 and HMX2 affecting the most genes with 54 and 23 genes affected respectively. We then explored if the variation of the level of expression of the cis genes were causally affecting the level of expression of the trans genes using Bayesian networks and found that differences in variation in the level of expression of the cis gene and variation of the level of expression of the trans gene. However, most of the trans associations (73.8%) are independent of the cis gene expression, implying that the trans associations are due to other functional variants in linkage with the cis-eQTLs or through other functional elements regulated in cis. This analysis shows that a large sample size allows the discovery of secondary effects of human variations on gene expression that can be used to construct short directed gene regulatory networks.

644W Common genetic variation within transcription factor binding sites is associated with bipolar disorder. D.T.W. Chen1, N. Akula1, L. Hou1, L. Jing1, G. Hawariat1, S. Detera-Wadleigh1, K. Jiang1, G.W. Black1, S. Dora-Wadleigh1, S. Doto1, W. Anderson1, P. Boulanger1, J. Cullinan2, D.M. Evans3, S. Detera-Wadleigh1, K. Jiang1, G.W. Black1, S. Doto1, W. Anderson1, P. Boulanger1, J. Cullinan2, D.M. Evans3, J.F. McMahon1. 1) Human Genetics Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 2) Department of Psychiatry, University of San Diego, La Jolla, CA 92039-0603.

Genome-wide association studies (GWAS) have uncovered a number of loci associated with bipolar disorder (BD), but functional alleles have not been identified. Some GWAS signals may reflect genetic variation in sequences regulating gene expression via binding of transcription factors (TFs). This study tested the hypothesis that signals identified via GWAS reflect genetic variation in transcription factor binding sites (TFBS). Genetic association data were extracted from a meta-analysis of worldwide GWAS in BD comprising ~14,000 cases/controls. This set of 700,000+ single nucleotide polymorphisms (SNPs) was mapped onto published TFBS identified in human lymphoblastoid cells. GWAS signal enrichment near TFBS was tested with the Kolmogorov-Smirnov (K-S) rank-sum statistic. Biological function of the TFBS was explored by use of the bioinformatics tools INRIICH, GREAT, and DAVID. Significant enrichment of GWAS signals (empirical p=9.1x10-3) was noted among SNPs near TFBS. This enrichment was not attributable to linkage disequilibrium between SNPs, or differing allele frequency spectra among SNPs detected in the GWAS. Genes nearest these SNPs significantly clustered into relatively few functional pathways. This clustering was not due to gene size or variable assignment of SNPs to nearby genes. Eight genes emerged most often (CAMKK2, ERN1, DDR1, MAP3K5, MAP4K5, MARK2, RAF1, BMP1A1); four of which encode proteins in the serine/threonine-protein kinase pathway. These findings suggest that regulation of gene expression by transcription factors plays an important role in the genetic etiology of bipolar disorder.
645T  
**Effect of Transcription Factor Binding Variation Depends on Genomic Context.**  
D.A. Cusanovich1, B. Pavlovic1,2, J.K. Pritchard1,2, Y. Gilad1.  
1 Human Genetics, University of Chicago, Chicago, IL; 2 Howard Hughes Medical Institute, University of Chicago, Chicago, IL.

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

646F  
**Localizing ancient causal regulatory variants from global genetic analyses of gene expression.** M.K. DeGorter1,2, S.B. Montgomery1,2.  
1 Department of Pathology, Stanford University, Stanford, CA; 2 Department of Genetics, Stanford University, Stanford, CA.

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

647W  
**Regulatory Function of CACNA1C Schizophrenia-Associated Variants.**  
N. Eckert1, R. Wang2, M. Szymanski-Pierce3, M. Zeleton1,2, S. Goswami2, D. Valle1, B. Auramopoulos1,2.  
1 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2 Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD.

Schizophrenia (SZ) is a complex psychiatric disorder affecting approximately 1% of the population with an estimated heritability between 70-90%. Genome wide association studies have identified many potential susceptibility variants for SZ, some of which have also been shown to be associated with Bipolar Disorder (BP). One such variant, rs1006737 in an intron of the CACNA1C gene, has been reported to be associated with both disorders. Two others in the same gene, rs7972947 and rs4765913, have been reported to be associated with SZ or BP. However, little is known about the identity of the causative variants and the mechanisms by which they contribute to pathogenesis. Because these variants are in non-coding regions, we hypothesized that they contribute to disease by disrupting normal regulation of gene expression. To test this hypothesis, we genotyped the three variants with Taqman assays and used RT-qPCR to measure expression of 3 alternative transcripts of CACNA1C in postmortem superior temporal gyrus (STG) brain samples from pathology-free controls. Then, we used generalized linear models to identify correlations between genotype and transcript levels. We found that the minor allele of rs1006737 is associated with decreased abundance of all CACNA1C transcripts in the STG. To identify the functional variant(s) underlying this effect, we made a series of dual luciferase constructs for all SNPs tagged by rs1006737 with an r2 > 0.80 and included approximately 1kb of genomic context in each case. Dual luciferase reporter assays in two cell lines showed that 4 of 12 genomic loci harbored CACNA1C tagged variants had statistically significant allele-specific effects on luciferase expression. Furthermore, electrophoretic mobility shift assays (EMSAs) showed that of the 6 SNPs included in these 4 constructs also have allelic differences in binding protein complexes from nuclear extracts of the same cell lines. From this, we conclude that rs1006737, the non-coding variant in CACNA1C that is associated with both SZ and BP, marks a regulatory haplotype that differentially affects the expression of multiple CACNA1C transcripts in the STG. The transcriptional activity in the haplotype raises many interesting questions. Further planned studies include identification of the proteins that bind the regulatory elements in an allelic specific manner and characterization of any complex chromatin interactions that mediate the observed effects.
649F
Identification of enhancer-promoter interactions in the mammalian genome. Y.-C. Hwang\(^1\), Q. Zheng\(^2\), C.-F. Lin\(^2\), V. Valladares\(^1\), B.D. Gregory\(^1\), L.S. Weng\(^1\), H. Zhou\(^1\), R.P. Groudine\(^1\), J.J. Weng\(^1\), M. Snyder\(^1\), S.B. Montgomery\(^1\), R.C. Wachtler\(^1\), L.-S. Wang\(^1\)
1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA; 3) Department of Computer Science, Stanford University, Stanford, CA; 4) Department of Pathology, Stanford University School of Medicine, Stanford, CA


651T
A comprehensive genomic landscape of NRSF binding in various cell types. P. Jain, FP. Bethin, GM. Cooper, RM. Myers. HudsonAlpha Institute for Biotechnology, Huntsville, AL

Neuron Restrictive Silencing Factor (NRSF/REST) is a master regulator of neuronal genes and is dysregulated in various neurodegenerative diseases and cancers. To more comprehensively understand the roles of NRSF, we have used ChiP-seq, RNA-seq, DNA Methylation (Reduced Representation for Bisulphite Sequencing), DNase I hypersensitivity, and histone modifications assays for genome wide mapping of regulatory profiles of NRSF and its cofactors within multiple cell lines. We also leveraged ENCODE data to systematically compare NRSF binding activity with binding of dozens of other transcription factors. We have identified key components of differential regulation of target genes by NRSF. We find that NRSF repression function is greatly influenced by other co-binding TFs and not restricted to its known cofactors like sim3a and CoREST. The data suggest that co-occupancy of other TFs with NRSF reduces NRSF repression activity, with transcriptional repression usually observed only in the absence of co-factor binding. Additionally, higher levels of DNA methylation at NRSF target genes coincide with lower expression. We also find that NRSF motif conservation is very high within binding sites near repressed genes. Our data suggest that while NRSF binds widely in human cells, only a small subset of target genes, that tend to also be methylated, have highly conserved motifs, and be depleted for binding of other TFs, are actually repressed by NRSF. This conclusion poses interesting questions as to the nature and consequences, or lack thereof, of most NRSF binding activity in human genomes.

652F
Distal co-regulated regions are crucial for human phenotypes. K.J. Karczewski\(^1\), A. Battle\(^2\), D. Knowles\(^3\), M. Snyder\(^1\), S.B. Montgomery\(^1\), Y. Sugahara\(^4\), L.-S. Wang\(^1\), M. Kamaya\(^1\), R. Kojima\(^1\), S. Giona\(^1\), B.D. Wachtler\(^1\), A. Battle\(^2\), D. Knowles\(^3\), M. Snyder\(^1\), S.B. Montgomery\(^1\)
1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA; 3) Department of Computer Science, Stanford University, Stanford, CA; 4) Department of Pathology, Stanford University School of Medicine, Stanford, CA

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

Our data suggest that long-range regulation is crucial for understanding the complexity of human phenotypes.

650W
Small introns of firefly luciferases: structural characterization and their insufficient substrates for splicing in CHO cells. M Ishii\(^1\), R. Kojima\(^2\), S. Fukuda\(^1\), T. Tani\(^1\), Y. Sakaguchi\(^1\), Y. Tanimura\(^1\), M. Nakashima\(^1\), M. Oyama\(^2\), 1) Biotechnology Laboratory, Department of Applied Chemistry, Faculty of Engineering, Kogakuin Univ, Hachioji, Tokyo 192-0015, Tokyo, Japan; 2) Enviromental Chemistry Laboratory, Faculty of Chemistry, Kogakuin University, Hachioji, Tokyo 192-0015, Tokyo, Japan; 3) The Firefly-Breeding Project at Hachioji, Tokyo 192-0015, Japan

Most eukaryotic genes contain segments of coding sequences (exons) interrupted by noncoding sequences (introns). Introns are removed through splicing, which is nearly universal in eukaryotes. However, the general function and evolutionarily importance of introns remains unclear. We cloned and sequenced the genomic and cDNA clones encoding luciferase in the Japanese firefly, Luciola cruciata. The luciferase genes contained six introns, and sequenced the genomic and cDNA clones encoding luciferase in the Genome. The luciferase gene (with 6 introns) and cDNAs (without introns) were inserted into mammalian expression vectors and transiently expressed in CHO cell. Western blot analysis and photoluminescence measurement indicated that the luciferase gene didn't express functional luciferase in the CHO cells, whereas cDNA did. RT-PCR analysis also showed that all of firefly introns didn't be spliced from pre-mRNA transcribed the luciferase gene. These results suggest that there are different mechanisms in splicing and splicing factors between insect and mammal.
653W
Regulatory motif centric validation, dissection, and construction of transcriptional enhancers. P. Kheradpour1, 2, J. Ernst1, 2, A. Melnikov2, P. Rogov2, L. Wang2, X. Zhang2, J. Alston2, T.S. Mikkelson2, M. Kellis2, 1) MIT Computer Science and Artificial Intelligence Lab, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) UCLA Department of Biological Chemistry, Los Angeles, CA.

Large scale studies have associated thousands of variants with phenotypes. However, due to genomic linkage structure, only a fraction of these variants likely play a causative role. While variants that may affect protein coding sequences can be computationally predicted, identifying variants with regulatory effect is much more difficult. To assess the feasibility of introducing variants and measuring the regulatory, we predicted causal regulators and employed a large-scale experimental approach to validate their instances. Using motif enrichments and depletions in ENCODE candidate enhancer regions, we predicted five activators (HNF1, HNF4, FOXA, GATA, NF-E2L2) and two repressors (GFI1, ZFP161) to be active in KS62 and HepG2 cell lines. We synthesized and performed a massively parallel reporter assay (MPRA) on 2,104 wild-type sequences centered on motif instances for these factors and an additional 3,314 engineered enhancer variants containing targeted motif disruptions.

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

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

655F

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

656W
Human Transcriptome Landscape Characterized by Deep RNA Sequencing in 957 Individuals. x. zhu1, 2, S. Mostafavi3, 4, A. Battle5, K. Beckman1, C. Hautenschild1, C. McCormick4, D. Koller5, AE. Urban1, 2, DF. Levinson1, SB. Montgomery2, 4, 6, 1) Psychiatry, Stanford University, Palo Alto, CA; 2) Computer Science Department, Stanford University, Palo Alto, CA; 3) Biomedical Genomics Center, University of Minnesota, Minneapolis, MN; 4) DNA sequencing services, DNA Sequencing Business, Illumina, Inc. Hayward, CA; 5) Department of Genetics, Stanford University, Palo Alto, CA; 6) Department of Pathology, Stanford University, Palo Alto, CA.

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

654T
Discovery and analysis of over 50,000 common functional regulatory variants in the human genome. M.T. Maurano, E. Haugen, E. Rynes, R. Humbert, J.A. Stamatoyannopoulos. Department of Genome Sciences, University of Washington, Seattle, WA.

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

Naive CD4+ T cells can differentiate into specific helper and regulatory T cell lineages in order to combat infection and disease. The correct response to cytokines and a controlled balance of these populations is critical for the immune system and the avoidance of autoimmune disorders. To investigate how early cell fate commitment is regulated, we generated the first human genome-wide maps of histone modifications that reveal enhancer elements after 72 hrs of in vitro polarization toward T helper-1 (Th1) and T helper-2 (Th2) cell lineages. Our analysis indicated that at this very early time point, cell-specific gene regulation and enhancers are at work directing lineage commitment, and likely opposed a role in cell maintenance as is differently regulated.

We determined a unique set of enhancer elements relative to recent ENCODE data in fully differentiated cells. Nucleosome-free regions were determined within enhancer chromatin structures. Examination of enhancers at nucleosome resolution identified transcription factor binding site (TFBS) motifs for expressed TFs with known and unknown T cell roles as putative drivers of lineage-specific gene expression.

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

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

Analysis of allele specific expression in mouse liver by RNA-Seq: marked differences compared to cis-eQTL identified using genetic linkage. P.-F. Roux1,2,3, S. Lagarrique2,3, L. Martin1, F. Hormozdarian4,5,6, A. van Nas6, O. Demeure2,3, A. Ghazalpour7, E. Eskin5,6, A.J. Lusis2,6, 1) INRA, UMR1348 Pegase, Rennes, France; 2) Agrocampus Ouest, UMR1348 Pegase, Rennes, France; 3) Université Européenne de Bretagne, France; 4) Department of Medicine/Division of Cardiology, University of California, Los Angeles, United States of America; 5) Department of Computer Sciences, University of California, Los Angeles, United States of America; 6) Department of Human Genetics, University of California, Los Angeles, United States of America; 7) Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, United States of America.

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

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

A large-scale transcriptome study in the Sardinian population. M. Pala1,2,6, M. Marongiu1,2, Z. Zappala1, A. Mulas4, R. Cusano2, F. Crobu2, F. Reiner4, R. Berutti3,6 M.G. Piras7, C. Jones3, D. Schlessinger2,6,1 A. Acasis1, A. Angius3, S. Sanna4, F. Cucca2,6, S. Montgomery1,2,6, 1) Pathology and Genetics Dept, Stanford University, Stanford, CA; 2) Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Monserrato, 09042, Italy; 3) CRS4, Advanced Genomic Computing Technology, Pula, Italy; 4) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 5) Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA; 6) Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy; 7) co-author.

The interpretation of genome-wide association studies (GWAS) is very challenging since most candidate loci fall in non-coding regions and are difficult to interpret. Given the important role of regulatory variation in phenotypic complexity, expression quantitative trait loci (eQTLs) have been proposed as an informative intermediate phenotype between genetic variation and human disease. To study the role of transcriptional mechanisms on regulating eQTL targets, we sequenced the polyA RNA fraction of blood mononuclear cells (PBMCs) isolated from 624 related individuals (259 families). These individuals had been subject to whole genome sequencing and characterized for more than 800 quantitative traits (including > 250 immune cell traits) by the Sardinia project. To assess the impact of eQTLs on lincRNAs, we also sequenced the RNA of a subset of 80 individuals after ribosomal RNA depletion. Our preliminary results show that the expression levels of 10,389 genes are associated with a genetic variant (top SNP with p-value ≤ 10−8 uncorrected), 40 of which are present in the GWAS catalog.

We also assessed splicing events and identified 27,904 donor splice sites associated with a genetic variant (top SNP with p-value ≤ 10−8 uncorrected), of which only 4 are present in the GWAS catalog. Using a subset of 68 individuals, we identified 9,427 heterozygous sites which show allele-specific expression (p-value ≤ 0.05) of which show high allelic imbalance (a mean of 0.80 for the most expressed allele) and low variability between individuals. These sites are enriched for missense variants and located near genes implicated in disease. In addition, we developed a statistical method to identify genes that are specifically and significantly over or underexpressed within single families compared to the majority of the population. We identified 39 genes that have family specific expression (p-value ≤ 0.01; FDR 0.462). Interestingly, these genes are also enriched for rare variants within 70kb upstream of their transcriptional start site, and we hypothesize that they may have a role in disease. We are now extending these analyses to the entire cohort and will present these analyses. Our goal is to identify associations between genotype and gene expression in order to enhance our overall understanding of gene regulation and correlate these molecular events with clinical data in order to characterize their role in phenotypic complexity and disease.
Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.

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

After a standard quality control, we obtained genotypes of 1,425,832 SNPs and 34,872 gene expression phenotypes from peripheral blood cells of 298 unrelated subjects recruited from the Japanese population. We carried out a genome-wide eQTL mapping for all the phenotypes. We identified 207,462 SNPs that were used in the analysis. Linear regression was performed to test associations between individual SNPs and probe-set level expression or gene level expression. An FDR value of 10% based on permutation was used to determine the significance level for cis probe-set level associations and gene level associations separately. As a result, 2503 cis gene level expression signals were identified. To further identify splicing signals controlled by SNPs, probe-set level associations where probe-set expression was highly correlated with gene level expression (r2>0.5) were excluded. As a result, 7941 genes with cis splicing signals were identified. Based on the patterns of alternative splicing, splicing signals were classified into categories such as exon skipping, alternative donors or acceptors, complex changes of multiple event types. To confirm the identified splicing signals, splicing signals targeted by each SNP was being performed on several candidate signals selected from different categories. To our knowledge, this is the first whole transcriptome array study to identify alternative splicing signals controlled by SNPs using human brain samples. As brain demonstrates a higher level of alternative splicing than any other tissue and a large number of neurological disorders have been linked to alternative splicing, the findings of the current study will help us understand genetic mechanisms underlying complex neurological disorders.

662W

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

Whole genome and exome sequencing data revealing that disease-associated genes harbor numerous completely benign missense mutations alongside pathogenic variants. Moreover, variants in some genes, such as LDDR1, are almost exclusively pathogenic, while other genes, such as Dicer1, contain mostly benign variants. Factors determining robustness of individual genes to sequence variations are unclear. We suggest that the robustness of genes is based on intrinsic structural properties of their encoded proteins, which we name ‘fragility’. We have constructed ‘fragility score’, which estimates the chance for a random amino-acid change to critically impact protein function, using DNA sequencing data on thousands of individuals from the general population from the Exome Sequencing Project (ESP) dataset and data on pathogenic alleles reported in the human gene mutation database (HGMD). High fragility score was assigned to proteins with significantly fewer missense mutations in the general population than in disease-associated populations, since most substitutions are deleterious. Using statistical and computational approaches, we estimated important evolutionary and functional predictors of pathogenicity, such as selective constraints and conservation of sequence across species. We analyzed hundreds of protein tertiary structures from the protein data bank to determine structural and thermodynamic determinants of protein fragility. Our results demonstrate that protein fragility derived from sequencing data is correlated to evolutionary conservation scores, to folding free energy change due to a missense mutation, and to the extent of intrinsically disordered regions within proteins. Characterization of fragility elucidates the biophysical constraints of protein structure and can inform various aspects of research on human genetic variation, ranging from basic population and evolutionary genetics, to genetic of complex traits, and clinical genetic diagnostics.

663T


The goal of this study was to use a novel, whole-transcriptome array to identify SNPs that affect alternative splicing in human brain. RNA was extracted from the prefrontal cortex of 371 samples (190 Alzheimer’s disease, 74 Huntington’s disease, and 107 controls) collected at the Harvard Brain Tissue Resource Center. A custom Affymetrix array with ~1 million probe sets specific to exons and junctions of 20,000 well-established human genes and un-annotated stretches of the genome was designed. This allows expression measurement at exon level in addition to gene level. Quality control on the new whole-transcriptome data was conducted using principal component analysis. Gene level expression was summarized by averaging expression of constitutive probe-sets (exon probe-sets that cover more than 50% of alternative transcripts). SNP genotyping was conducted on an Illumina 650K array. After quality control and imputation, a total of 11.5 million SNPs were used in the analysis. Linear regression was performed to test associations between individual SNPs and probe-set level expression or gene level expression. An FDR value of 10% based on permutation was used to determine the significance level for cis probe-set level associations and gene level associations separately. As a result, 2503 cis gene level expression signals were identified. To further identify splicing signals controlled by SNPs, probe-set level associations where probe-set expression was highly correlated with gene level expression (r2>0.5) were excluded. As a result, 7941 genes with cis splicing signals were identified. Based on the patterns of alternative splicing, splicing signals were classified into categories such as exon skipping, alternative donors or acceptors, complex changes of multiple event types. To confirm the identified splicing signals, splicing signals targeted by each SNP was being performed on several candidate signals selected from different categories. To our knowledge, this is the first whole transcriptome array study to identify alternative splicing signals controlled by SNPs using human brain samples. As brain demonstrates a higher level of alternative splicing than any other tissue and a large number of neurological disorders have been linked in aberrant splicing, the findings of the current study will help us understand genetic mechanisms underlying complex neurological disorders.
Gene expression profiling of young and adult mouse cochlea by RNA-Seq in strains with normal and age-related hearing loss. A.B. Giersch1, J. Shen1, N.G. Robertson2, K. Wong3, C.C. Morton1,2,3. 1) Pathology, Brigham and Women’s Hospital, Boston, MA, USA; 2) Obstetrics and Gynecology, Brigham and Women’s Hospital, Boston MA, USA; 3) Harvard Medical School, Boston, MA, USA

Purpose: Age-related hearing loss is the most common sensory deficit in humans, reducing the quality of life in the aged population. Its social impact will become more pronounced as life expectancy continues to increase. Despite the discovery of many deafness genes, understanding of the pathophysiologic mechanisms underlying age-related hearing loss remains elusive. We hypothesize that gene expression profiling of the cochlea from mouse models with various degrees of age-related hearing loss will reveal the molecular mechanism and inform potential target selection for prevention. Method: We performed gene expression profiling of mouse cochlea by next-generation sequencing (RNA-Seq). A multi-factorial design was used. Cochleas from mouse strains with good hearing past one year of age (CBA/Caj and B6.CAST-Cdh23Ahi+) or with documented age-related hearing loss (CochG868G/E868E and Coch--) in a CBA background, and C57BL/6J were dissected at discrete ages ranging from one week through late adulthood. Poly(A) selected mRNAs were extracted and the derived cDNA samples were fragmented, indexed, pooled, and sequenced by Illumina HiSeq. Biological replicates were used for all conditions. Expression levels of all transcripts were analyzed and differential gene expression analyses were performed. Results: We have obtained gene expression profiles of mammalian cochlea at various ages by RNA-Seq. With total reads of at least 30 million, more than 16,000 genes were detectable in each sample. The expression levels were highly reproducible. We detected significant systematic differences in gene expression profiles between C57BL/6J and CBA/Caj strain backgrounds, regardless of age. Comparing mouse models with or without age-related hearing loss of the same genetic background, we have found that few genes show significant differential expression at young ages before the onset of hearing impairment, but the number of genes dramatically increases to hundreds at later stages. In addition, hundreds of genes show significant temporal changes on both genetic backgrounds. Summary: We have performed gene expression profiling of human colon biopsies each are treated with 0.1M 1,25(OH)2 vitD or vehicle control (EtOH). We found 508 of 11,317 differentially expressed genes in response to vitD treatment (FDR< 0.01). Among 307 up-regulated genes, we found several known vitD responsive genes including CYP24A1 (p=2.6x10-16), TRPV6 (p=6.93x10-16), and CD14 (p=2.2x10-16). In addition to CD14, we found a number of up-regulated immune-related genes including IRF8 and TLR4. In a preliminary study of 24 additional individuals, we have replicated these findings. We also noted that for 33 genes there were significant differences in genome-wide log-fold change between AA and EA. Among these, suppressor of AP-1, regulated by IFN (SARI or BAF72) was significantly upregulated in response to vitD in EAs but not in AAs. In summary, we found significant up- and down-regulated genes in response to 1,25(OH)2 vitD in human colon biopsies. Several known vitD-responsive genes were up-regulated. We noted several up-regulated immune-related genes including CD14 and TLR4. We also found genes with significantly different log-fold change in response to vitD between AA and EA suggesting that inter-ethnic differences in the vitD pathway include transcriptional response to a given amount of vitD. Further studies of genes identified in this study may reveal important vitD responsive pathways that underlie disease disparities.
667F  
Gene Expression Studies by RNA-Seq in Airway Epithelial Cells (AECs) from Asthmatic and Non-Asthmatic Individuals. R.A. Myers1, J. Nicodemus-Johnson1, D.K. Hogarth2, J. Sud4, J.P. McConville1, E.T. Naureckas3, A.I. Sperling1, J. Solway3, J.A. Krishnan3, S.R. White4, D.L. Nicolae1,2,3, C. Ober4. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL. 

Asthma is a chronic inflammatory disease characterized by reversible airflow obstruction. Airway epithelial cells (AECs) form the interface between the environment and the host and are functionally important in asthma pathogenesis. We hypothesized that transcript abundance in AECs differs between asthmatic and non-asthmatic individuals. In this study, we characterized the transcripts present and identify differentially expressed genes by RNA-Seq in 48 airway epithelial cell samples (27 asthmatics, 21 non-asthmatics) using the Illumina HiSeq 2000 platform. Sequences were mapped to the genome using BWA and the number of sequences mapped to protein coding genes was determined using BEDTools. We used edgeR to test for differential expression controlling for technical and biological factors that influence gene expression. We identified as expressed 14,484 genes with at least one count per million sequences (cpm) in at least nine samples. At a false discovery rate of 10%, 43 genes were differentially expressed, 25 with increased expression and 18 with decreased expression in asthmatics compared to non-asthmatics. The gene with the most significant differential expression is regulator of G-protein signaling 2 (RGS2, p-value = 5.39 x10-9), which showed increased expression in asthmatics. RSG2 inhibits G-protein signal transduction by increasing GTPase activity of G-protein alpha subunits. As expression of RGS2 is induced by long acting beta2-adrenoceptor agonists and glucocorticoids, the gene expression pattern may reflect responses to medication in the individuals with asthma. Other differentially expressed genes included ALOX15B, involved in arachidonic acid metabolism; TC1 (C8orf4), a regulator of inflammation in arachidonic acid metabolism; CCR2; and KCNJ5, a subunit of a channel involved in neuronal response; chemokine receptors CCR2 and CX3CR1; and KCNV5, a subunit of a G-protein activated potassium channel. Ten pathways showed enrichment (p < 0.01) of differentially expressed genes (p-values < 0.05 and fold changes > 2), including the antigen presentation pathway (p-value = 0.0005), antigen processing and presentation pathway (p-value = 0.0027). Through RNA sequencing in AECs, we identified differentially expressed genes and pathways that may be potential targets for asthma treatment. This work is supported by NIH grant U19 AI095320.

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

EQTLs and allele specific expression of HLA haplotypes and amino acids associated to autoimmune diseases. A. Zhernakova1, M. vd Sijde2, J. Gutierrez-Achury1, P.J. McLaren3, D.V. Zhernakova4, P.I.W. de Bakker2,4, H.J. Westra1, L. Franke1, C. Wijmenga1, J. Fu1. 1) Genetics, UMC, Groningen, Netherlands; 2) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Switzerland; 3) Department of Medical Genetics, and Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Division of Genetics, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts, USA.

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

Host and pathogen transcriptome profiling during Leishmaniasis major infection. L.A.L. Dillon1,2, R. Suresh1, K. Okrah2, M. Mangione3, J. Choi4, H. Corrada Bravo5,6, D.M. Mosser1, N.M. El-Sayed1,2. 1) Department of Cell Biology & Molecular Genetics, University of Maryland, College Park, MD; 2) Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD; 3) Department of Mathematics, University of Maryland, College Park, MD; 4) Department of Computer Science, University of Maryland, College Park, MD.

Leishmaniasis, caused by protozoan parasites of the Leishmania genus, affects roughly 1.5 million people worldwide and is endemic primarily to South America and the Middle East. The parasite’s lifecycle is divided between its insect vector, the phlebotomine sand fly, and its mammalian host, where it resides inside of macrophages. In order to establish an infection, the parasite must evade host immune system responses to infect macrophages and to survive and replicate within cellular phagolysosomes. Little is known about changes that take place at the transcriptional level in either the host or the pathogen during an infection and how Leishmania is able to direct these changes through manipulation of host cellular machinery. Using the mouse as a model system, we have performed RNA-seq on the illumina HiSeq platform to simultaneously measure global changes in gene expression in both L. major and host macrophages at multiple time points during the L. major lifecycle, including as the parasite replicates inside of host cells. Transcripts were aligned to the mouse and L. major genomes using TopHat and differential expression analysis was performed using limma after ComBat batch correction and voom data transformation. Gene ontology category enrichment analysis was done using GOSeq. Genes that are differentially regulated in L. major include histones, antioxidant genes, cell surface markers, membrane transporters, and signaling molecules while those that are differentially expressed in mouse include those involved in immune system responses (e.g., inflammation, wound healing, leukocyte activation), cell signaling, metabolism, and apoptosis.

The identification and quantification of up- and down-regulated genes has provided evidence regarding the mechanisms used by L. major to elude host defenses and survive in the intracellular environment and has enabled connections to be made between changes in gene expression and the parasite’s biology and lifecycle. These findings are revealing new insights into the dynamics of an infection and will help identify markers and targets for the prevention, diagnosis, and treatment of Leishmaniasis and related kinetoplastid diseases.
Peripheral blood microarrays identify dysregulated pathways and pathways unique to psoriatic arthritis compared to psoriasis without arthritis. R. Postolaki\(^1\), V. Chang\(^1\), T. C. Virtanen\(^1\), F. Pellet\(^1\), C. Rosen\(^2\), H. Liang\(^2\), D. Gladman\(^1,2\). 1) Psoriatic Arthritis Program, Toronto Western Research Institute, Toronto, Canada; 2) Division of Rheumatology, Department of Medicine, University of Toronto, Toronto, Canada; 3) Ontario Cancer Institute Genomics Centre, Toronto, Canada; 4) Division of Dermatology, Department of Medicine, University of Toronto, Toronto, Canada; 5) Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada.

Psoriatic arthritis (PsA) is an inflammatory arthritis that develops in about 30% of patients with psoriasis (PsC). We aimed to identify a gene expression signature of PsA compared to PsC and controls to identify biomarkers and improve our understanding of the pathology of arthritis in psoriasis patients. Gene expression was measured in peripheral whole blood from 20 PsA patients, 20 PsC patients without arthritis, and 20 controls using Agilent Whole Human Genome Oligo microarrays. Statistical analyses were performed using commercial (GeneSpring GX) and open-source (Bioconductor) packages to identify significantly differentially expressed genes (DEGs) between PsA vs. controls, PsC vs. controls, and PsA vs. PsC. DAVID and Ingenuity Pathway Analysis were performed to identify enriched annotations and pathways relative to the Agilent probe set reference. Eleven candidate biomarker genes were validated by qPCR using Taqman assays. Expression levels across all three groups were compared using one-way ANOVA and 1547 DEGs were found. Student's t-test found 790 DEGs between PsA vs. controls, 98 between PsA vs. PsC, and 0 between PsC vs. controls. DEGs in PsA vs. controls included genes involved in peptide loading onto the MHC class I complex, DNA damage and DNA repair, cell adhesion, and cell cycle, as well as genes involved in protein synthesis and protein translation. Activation of mature and precursor forms of the transmembrane protein hippo (STRA) and histone acetyltransferase complexes (CSTA), histone acetyltransferase complexes (EP300, CREBBP, MYST4), toll-like receptor 2 signaling (AKAP13), and NKT cell migration (S1PR5). Two-way hierarchical clustering of the DEGs with a fold change > 2 grouped 18 out of 20 PsA patients together, with the remaining 2 patients clustering with PsC patients. Statistically significant pathways related to gene expression in PsA vs. PsC were identified using DAVID. Deep Sequencing of microRNAs and novel tRF RNAs in human monocytes. C. Escobar\(^1\), H. Ge\(^1\), C. Li\(^2\), C. Rosen\(^2\), H. Ameti\(^2\), T. Juan\(^2\), R. Sandrock\(^\dagger\). 1) Genome Analysis Unit, Amgen Inc., San Francisco, California, 94080; 2) Protein Sciences, Amgen Inc., Thousand Oaks, California, 91320; 3) Inflammation Research, Amgen Inc., San Francisco, California, 94080; 4) Protein Sciences, Amgen Inc., Thousand Oaks, California, 91320; 5) Inflammation Research, Amgen Inc., Thousand Oaks, California, 91320. MicroRNAs play essential roles in modulating the level of gene expression post-transcriptionally in cells. Certain microRNAs have been shown to be up- or down-regulated during the process of immune responses. Using next-generation sequencing (NGS), we performed transcriptional analysis of small RNAs ranging from 18- to 30-nt in unstimulated human monocytes or those treated with an activating stimulus, lipopolysaccharide (LPS). Upon aligning and counting fragments that mapped to the human genome, we found that 20 microRNAs accounted for approximately 75% of the entire microRNA population in monocytes. Within around 5-order range of differential expression, miR-155 was the most abundant up-regulated gene in LPS-induced monocytes followed by miR-146a, miR-155, miR-15b, and miR-34. No significant down-regulation of microRNAs was observed in human LPS-treated monocytes. Activation of mature and precursor forms of the signature set of microRNAs was further confirmed in LPS-treated monocytes by a semi-quantitative analysis of tagged RT-PCR. In addition to known microRNAs, we identified 18 IRF RNA fragments (IRFs) sequences, an emerging type of novel small RNAs, of which little biological significance is known, that were originally identified in cancer cells that mapped to 50 different IRF gene loci of which 13 were novel. IRFs are a type of novel small RNAs, of which little biological significance is known, that were originally identified in cancer cells. Only RNA-Tyr7G showed differential expression in human LPS-treated monocytes. Our study provides a comprehensive quantitative catalog of microRNA expression and demonstrates an analysis pipeline for identification of novel IRF RNAs that lays the groundwork for further investigation regarding the biological significance of IRFs in human monocyes.
675T Platelet RNA And eXpression-1 Study Demonstrates RNA Expression Differences that Correlate with Ancestry. L. Simon1, E. Chen1, L. Edelstein1, P. Bray2, C. Shaw1. 1 Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2 Thomas Jefferson University, The Cardeza Foundation for Hematologic Research and the Department of Medicine, Jefferson Medical College, Philadelphia, PA, USA.

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

676F Genetic variation underlying protein expression variation in brain samples and lymphoblastoid cell lines. R.J. Hause1,2,3, C.T. Archer4, L. Cheng2, J.A. Badner1, C. Liu1, R.B. Jones1,2,3. 1 Ben May Department for Cancer Research, University of Chicago, Chicago, IL; 2 Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL; 3 Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL; 4 Department of Human Genetics, University of Chicago, Chicago, IL; 5 Department of Psychiatry, University of Illinois at Chicago, Chicago, IL; 6 Department of Psychiatry and Behavior Neuroscience, University of Chicago, Chicago, IL.

Genome-wide association analyses have discovered many DNA variants that influence complex psychiatric disorders. Because gene expression is a molecular phenotype that acts as an ‘intermediary’ between genetic and physiological variation, expression quantitative trait loci (eQTL) mapping has been used to try to better understand complex disease by identifying regulatory variation that affects mRNA expression levels. An implicit assumption made in these analyses is that eQTLs explain subsequent differences in protein levels, even though previous experiments have shown imperfect correlations between mRNA and protein expression levels. To improve our understanding of the genetic basis of protein expression variation and how this variation is related to genetic risks for complex disease, we have recently extended micro-western and reverse phase lysate array technologies to quantify 441 protein levels across 68 unrelated Yoruba lymphoblastoid cell lines (LCLs) and examined the correlations between mRNA levels, protein levels, and known physiological covariates. In this study, we sought to translate our protein analysis work to examine protein levels in the brain.

We quantified the levels of 50 proteins related to cell signaling, transcription, and neurobiology across brain samples derived from a population of 129 unrelated individuals of Caucasian descent who were classified as normal controls (n = 43) or diagnosed as cases with depression (n = 12), schizophrenia (n = 39), or bipolar disorder (n = 35). We first identified several proteins that were differentially expressed between cases and controls, such as brain-derived neurotrophic factor (BDNF), and between particular psychiatric sub-classes and controls, such as thymus-specific serine protease (PRSS16) that was significantly overexpressed in depressed patients. We then performed association analyses on these protein levels to identify protein quantitative trait loci (pQTLs). We observed that up to a quarter of pQTLs previously identified in LCLs replicated in this brain cohort. Our approach represents a general method for the identification of novel protein variation associated with neuropsychiatric disorders and genetic variation that contributes to protein variation. In the future, we will investigate the overlap of pQTLs with SNPs previously implicated in neuropsychiatric disorders and scale up our investigation to analyze a substantially larger protein set.

677W Identification of novel non-coding RNAs associated with smoking. M.J. Peters1,2, F.A.S. Smouter1,2, J.G.J. van Rooij1,2, D. Schmitz1, M. Jhamai1, P. Arp2, R. Rivadeneira1,2,3 A. Hofman2, A.G. Ulteerlinden1,2,3, J.B.J. van Meurs1,2, 1 Department of Internal Medicine, Erasmus Medical Center Rotterdam, the Netherlands; 2 The Netherlands Genomics Initiative-sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA), Rotterdam, the Netherlands; 3 Department of Epidemiology, Erasmus Medical Center Rotterdam, the Netherlands.

Background Smoking is the leading cause of premature death from diseases such as lung cancer and chronic respiratory disease; it harms nearly every organ and reduces the health of smokers in general. Since the toxic ingredients in cigarette smoke (like nicotine) are absorbed into the bloodstream, we hypothesized that smoking will alter the gene expression patterns of many coding and non-coding RNAs in the circulation. Our aim was to identify these RNAs changing with smoking. Oxidative stress is thought to be the general cause: smoking causes an imbalance between the systemic manifestation of reactive oxygen species and the system’s ability to repair the resulting damage. Methods We performed RNA sequencing in 92 women of the Rotterdam Study: 30 smokers and 62 non-smokers. Whole-blood was collected using PAXgene-tubes and the RNA was sequenced using the Illumina HiSeq 2000 (4 samples/lane). We aligned the reads to the Human Reference Genome (v19) using TopHat and we estimated the relative abundance using Cufflinks (Tuxedo Suite); transcripts with a coverage <3 were removed and they needed to be present in >2 samples. We used logistic regression to model the outcome variables and adjusted for age, RNA quality, technical batch, and cell counts. We mapped the transcripts to different RNA reference databases to distinguish known and novel coding and non-coding RNAs. Results We identified a total of 112 RNAs differentially regulated between smokers and non-smokers (p<0.01), of which 71 were present in any RNA reference database and 41 were novel. Examples of transcripts known to be associated with smoking behavior are ANKRD13B, GPRF2, UBOULNL, PIGQ and OLFM1. We newly identified the TXNRD3 transcript: four smokers had transcribed this gene, while none of the non-smokers had detectable TXNRD3 RNA levels. The TXNRD3 protein catalyzes the reduction of thioredoxin, and is implicated in the defense against oxidative stress. 41 novel RNAs were differentially regulated between smokers and non-smokers: most interesting are the ones present in >=10 non-smokers and not present in smokers at all. These new intergenic transcripts are located upstream of the LEP1 gene and upstream of the CUFF1 gene. Conclusions These results show that smoking can alter the expression of novel RNAs, which can offer new targets for intervention or as biomarkers of smoking behavior. While replication is pursued, identification of both known and novel RNAs associated with smoking will add to our understanding of the molecular response in blood to cigarette smoke exposure.
Using Metabolomics and ex vivo Activity Approaches to Understand the Functional role of Epoxide Hydrolyase 2 gene in Anorexia Nervosa. P. Shih1, C. Morisseau2, J. Yang3, T. Clarke3, A. Van Zeeland4, A. Bergen4, P. Magistretti5, N. Schork6, W. Benrrettini5, B.D. Hammock7, W. Kaye1, Price Foundation Collaborative Group. 1) Dep of Psychiatry, Univ California, San Diego, La Jolla, CA; 2) Dep of Endotomology, Univ California, Davis, Davis, CA; 3) Dep of Psychiatry, Univ of Pennsylvania, Phila., PA; 4) The Scripps Translational Science Institute, La Jolla, CA; 5) SRI International, Menlo Park, CA; 6) The University of Lausanne, Lausanne, Switzerland.

Individuals with Anorexia nervosa (AN) restrict eating and become emaciated. They tend to have an aversion to foods rich in fat. Understanding the genetic basis of fat metabolism may unravel AN molecular psychopathology. We recently identified a novel AN susceptibility gene, Epoxide Hydrolyase 2 (EPHX2), through a series of complementary genetic study designs (GWAS, exon-sequencing, single-locus association and replication studies) in 1205 AN and 1948 controls (p=0.0004 to 0.0000016). The molecular mechanism by which EPHX2 influences AN risk remains elusive; here we utilize metabolomics and ex vivo enzyme activity assays to evaluate the biological function of EPHX2. EPHX2 codes for soluble epoxide hydrolase (sEH) which hydrolizes and converts derivatives of arachidonic acid (AA), Epoxyeicosatrienoic acids (EETs) to Dihydroxyeicosatrienoic acids (DHTEs). We measured AA metabolite markers (8,9-, 11,12- and 14,15-EETs) and their corresponding DHTEs ratios (p=0.04 to 0.0016), suggesting elevated basal sEH activity in AN and controls except for increased 8,9- and 14,15-EETs in AN (p=0.05). Variant allele carriers of Arg287Gln and 3’-UTR SNP (rs1042064) in EPHX2 showed decreased levels of all 3 EETs (p=0.04 to 0.0002) and lower 11,12- and 14,15- EET-to-DHTEs ratios (p=0.04 to 0.0016), suggesting elevated basal sEH activity in variant allele carriers. However, the ex vivo measurement of sEH activity did not show association with genotypes. Our earlier work using the bacculovirus surrogate substrate on buffy coat in an independent sample of 36 controls. No significant differences in AA metabolite markers were observed between controls with Arg287Gln (rs751141) and 3’-UTR SNP (rs1042064) genotypes. Our earlier work using the bacculovirus surrogate substrate on buffy coat in an independent sample of 36 controls. No significant differences in AA metabolite markers were observed between controls with Arg287Gln (rs751141) and 3’-UTR SNP (rs1042064) genotypes.

The African Genome Variation Project Phase II: Down-sampling of African whole genome sequence data. T. Carstensen1, D. Gurdasani2,3, N.S. Schork4, J. Yang3, T. Clarke3, A. Van Zeeland4, A. Bergen4, P. Magistretti5, N. Schork6, W. Benrrettini5, B.D. Hammock7, W. Kaye1, Price Foundation Collaborative Group. 1) Dep of Psychiatry, Univ California, San Diego, La Jolla, CA; 2) Dep of Endotomology, Univ California, Davis, Davis, CA; 3) Dep of Psychiatry, Univ of Pennsylvania, Phila., PA; 4) The Scripps Translational Science Institute, La Jolla, CA; 5) SRI International, Menlo Park, CA; 6) The University of Lausanne, Lausanne, Switzerland.

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

SRD5A2 gene polymorphisms affect the risk of benign prostatic hyperplasia. V.K. Choubey1, S.N. Sankhwar1, J. Carlus2, A.N. Singh3, D. Dalela3, R. Singh4. 1) Department of Urology, King George's Medical University, Lucknow, Lucknow, India; 2) Centre for Genetics and Inherited Diseases (CGID), Taibah University, Al-Madinah, Kingdom of Saudi Arabia; 3) Division of Endocrinology, Central Drug Research Institute, Lucknow, India.

Abstract Background Benign prostatic hyperplasia (BPH), characterized by an enlarged prostate, affects the quality of life in the elderly people, and prostate cancer (PC) is the second leading cause of death in men. Since prostate is an androgen dependent tissue, androgen metabolism is likely to affect the risk of developing BPH and PC. Methods We have analyzed common polymorphisms in the SRD5A2 gene in 210 BPH patients, 192 PC cases and 171 controls. Published data on V89L and (TA)n repeat polymorphisms was pooled for performing meta-analysis to quantitate the BPH risk associated with these polymorphisms. In total, six studies on V89L polymorphism comprising a total of 787 patients and 586 controls and three studies on (TA)n polymorphism comprising a total of 768 patients and 460 controls, were included in meta-analysis. Results We did not observe any polymorphism at the A49T polymorphic site. All the subjects showed ‘AA’ at this site without any substitution. The presence of ‘V’ increased BPH risk (p = 0.047); however, no significant difference in the genotype distributions comparison between PC vs Controls was observed. Comparison of genotypes at (TA)n locus found highly significant difference between cases and controls such that shorter repeats increased BPH risk (p = 0.003). However, no significant difference in comparison of PC vs Controls was observed. Meta-analysis also showed high frequency of ‘V’ at V89L and (TA)n repeats in the BPH cases, though the differences were not statistically significant. Conclusion A49T locus of SRD5A2 gene is monomorphic in the BPH cases, though the differences were not statistically significant. Conclusion A49T locus of SRD5A2 gene is monomorphic in the BPH cases. However, the presence of VV at V89L Locus and shorter (TA)n repeats in the BPH cases, though the differences were not statistically significant. Conclusion A49T locus of SRD5A2 gene is monomorphic in the BPH cases.
High-resolution personal genome-wide maps of meiotic double-strand breaks in humans. F. Pratto1, K. Brick1, P. Khil1, F. Smagulova1, V. Petkho 2, R.D. Camerini-Otero3. 1) Genetics and Biochemistry Branch, National Institutes of Health, Bethesda, MD, USA; 2) Department of Biochemistry and Molecular Biology, Uniformed Services University of Health Sciences, Bethesda, MD, USA.

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

Alternative promoter activation leads to the expression of a novel variant of human lysyl oxidase (LOX-v2) that functions as an amine oxidase.

Lysyl oxidase (LOX) is a copper-dependent amine oxidase that is responsible for the lysine-mediated cross-links found in the extracellular matrix proteins, such as collagen and elastin. Four additional LOX-like genes (LOX1, LOXL2, LOXL3, and LOXL4) have been identified in human, each encoding a 188 amino acid-long polypeptide of 22 kDa. In peroxidase-coupled fluorometric assays, LOX-v2 showed a significant amine oxidase activity toward lysyl-tyrosyl quinone (LTQ) and a cytochrome receptor-like (CRL) domain. Several novel functions including tumor suppression, tumor progression, cellular senescence, and chemotaxis have been recently attributed to LOX. The presence of LOX paralogues, thus, suggests that the diverse multiple functions of LOX may be derived from differential regulation of the LOX paralogues. In searches of more human LOX paralogues, we identified several expressed sequence tag (EST)-clones that showed an alternative exon-intron splice pattern from LOX. Those ESTs corresponded to the LOX transcript variant 2 (LOX-v2) that was recently reported in the GenBank (accession no. NM_00117812). LOX-v2 is lack of exon 1 of LOX, encoding a 188 amino acid-long polypeptide of 22 kDa. In peroxidase-coupled fluorometric assays, LOX-v2 showed a significant amine oxidase activity toward collagen and elastin. RT-PCR analysis with human tissues, LOX-v2 showed distinct tissue specificity from LOX. An alternative promoter element present in the intron 1 region of LOX was sufficient for the differential transcriptional activation of LOX-v2. These findings indicate that the human LOX gene encodes two variants, LOX and LOX-v2, both of which function as amine oxidases with distinct tissue specificities from one another.

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

Coeliac disease (CD) is a common, complex and chronic immune-mediated disease affecting the small intestine. It is triggered in genetically sensitive individuals by the ingestion of gluten proteins, which initiate an immune reaction that leads ultimately to the destruction of the normal architecture of the gut wall, most obviously the loss of the villus architecture and the development of crypt hyperplasia. There is a strong inherited component to the disease, demonstrated by the fact that concordance rates in monozygotic twins are at least 75%. HLA-DQA1 and HLA-DQB1 alleles are necessary but not sufficient to cause disease and recent genome-wide association studies (GWAS) have succeeded in identifying almost 40 non-HLA risk loci. For the majority of these non-HLA loci very little is known with regards their potential role in immune function and/or the mechanism by which they might be contributing to disease biology. Given that the activation of T cells in the intestinal mucosa in response to gluten exposure is thought to play a key role in the pathogenesis of the disease, we measured the mRNA levels of disease associated candidate genes specifically in activated and resting CD4+ T cells in a group of coeliac individuals and unaffected controls. Results indicate that a number of genes that have been genetically associated with the disease also show significantly different expression at the mRNA level in CD4+ T cells and further investigation of these genes through the integration of Immunochip derived genetic data and pathway analysis will hopefully lead to a greater understanding of their role in disease biology.
Real-time PCR for Indel markers detection in cffDNA. M.D. Santos, F.M.F. Nunes, A.L. Simões. 1) Department of Genetics, University of Sao Paulo, Ribeirão Preto, Sao Paulo, Brazil; 2) Department of Genetics and Evolution, Federal University of Sao Carlos, Sao Carlos, Sao Paulo, Brazil.

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


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

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

MHC gene family is comprised of three classes of genes, which are instrumental in determining donor-recipient compatibility for organ transplant as well as susceptibility to autoimmune diseases via cross-reacting immunization. Specifically Class I HLA-A, -B, -C, and Class II HLA-DR, -DQ and -DP genes are considered medically important for genetic investigation to determine histocompatibility. Highly polymorphic in nature, thousands of alleles of these HLA genes have been implicated in disease resistance and susceptibility, and the importance of full-length HLA gene sequencing for genotyping, detection of null alleles and phasing is being widely acknowledged. Although DNA-sequencing-based HLA genotyping is now routinely done, only 7% of the HLA genes have been characterized by allele-level sequencing, while 93% are still defined by partial sequences only. The gold-standard sequencing technology, though widely used, is unable to generate phased reads from heterozygous alleles. Second-generation, high-throughput clonal sequencing methods, although better at heterozygous allele detection, are inadequate at generating full-length haploid gene sequences from enhancer promoter to 3′UTR along with phasing information. The best way to overcome these challenges is to sequence these genes with a technology that is clonal in nature and has the longest possible read lengths. We have employed Single Molecule Real-Time (SMRT®) sequencing technology from Pacific Biosciences for sequencing full-length HLA class I and II genes. PCR systems were developed to amplify entire HLA genes ranging between 4.6 kb and 9.7 kb. For HLA genes that were too long, like HLA-DBR1 and -DPB1 (13–21 kb), the PCR regions were divided into two parts (enhancer-promoter to exon 2, and exon 2 to 3′UTR for HLA-DBR1; and enhancer-promoter to intron 2, and intron 1 to 3′UTR for HLA-DPB1). The long-range PCR amplicons were converted into SMRTbell™ libraries, and long reads ranging from ~3,500 to 20,000 bases were generated on the PacBio® RS. Full-length consensus reads were obtained with correct phasing information and accuracies for all the alleles of each of the HLA genes based on the zygosity and presence of pseudo genes. The resolving power of SMRT Sequencing for simultaneously genotyping, determination of phasing information, and detection of novel alleles makes it a unique multigene DNA-based method for unambiguous allele-level characterization of HLA genes.
689W
The Role Of MDR1 C3435T and C1236T Single Nucleotide Polymorphisms in Male Infertility. S. Aydos1, A. Karadag2, T. Ozkan3, B. Altink1, M. Bunsuz1, S. Heidarhizadeh1, K. Aydos1, A. Sunguroglu1. 1) Ankara University, Faculty of Medicine, Department of Medical Biology Ankara, Turkey; 2) Ankara University, Faculty of Medicine, Reproductive Health Research Center Ankara, Turkey.

Infertility is a common problem among couples around the world, affecting one in six couples. In 30% of infertile couples, the male factor is a major cause due to defective sperm quality. However, the factors responsible for defective sperm remain largely unknown. The multi-drug resistance 1 (MDR1) gene encoding a P-glycoprotein, which has a role in active transport of various substrates, including xenobiotics, and thus has a protective function in various tissues and organs. In the present case-control study, we investigated the effect of MDR1 gene C3435T and C1236T SNPs and on male infertility in Turkish population. The study was conducted on patients with infertility and 100 healthy control. The genotyping of C3435T and C1236T SNPs was done by PCR-RFLP. In statistical analysis were tested with chisquare test and SHEsis program. 101 patients were compared with sperm parameters. Frequencies of the C and T alleles of the 1236 locus were found to be 61.2% (n=235) and 38.8% (n=149) in the infertile patient, similar to the 59.8 and 40.2% found in the healthy control, respectively. The C allele was detected more frequently in the patient. The frequencies of MDR1 1236 CC, CT and TT genotypes were 34.4%, 53.6% and 12% in the patient; those in the control were 34.3%, 51%, 14.7%, consecutively. However, the observed genotype frequencies did not show significant difference in either group (P>0.05). Frequencies of the C and T alleles of the 3435 locus were found to be 56.2, 43.8 % in the infertile patient, similar to the 51%, 49% found in the healthy controls. The genotype frequencies of C3435 T and C1236 T genotypes were 27.6%, 57.3%, 15.1% in the patient; those in the control were 26.5%, 49%, 24.5%, consecutively. However, the observed allele and genotype frequencies did not show significant difference in either group (P>0.05). Each of the four possible haplotypes was noted in both infertile patient and control. When the frequency distributions of estimated haplotypes were compared between the patient and control, the frequency of the T-T haplotype was found to be significantly higher in the control than in infertile patients (P=0.006). There was no correlation between sperm parameters and genotypes. Our findings show that two SNPs do not play a role in the genetic susceptibility to male infertility, controversial to Polish population. But T-T haplotypes may be a protective factor for the fertility, because this haplotype was found to be statistically lower in infertile patients than in control.

690T
Interactions between Epigenetic and Genetic Signatures in the Asthmatic Airway. J. Nicodemus-Johnson1, R.A. Myers1, D.K. Hogarth1, J. Sudd1, J.F. McConville1, E.T. Naureckas1, J. Selvidge1, J.A. Krishnan1, S.R. White1, D.L. Nicol12, Y. Gilad1, C. Ober1. 1) Human Genetics, University of Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL. A large number of genes are differentially expressed in airway epithelial cells of asthma compared to non-asthmatic cells. The mechanisms underlying these differences are not well understood, and the role of genetic and epigenetic factors in these differences is yet to be fully elucidated. In this study, we aimed to investigate the interactions between genetic and epigenetic mechanisms using asthma-risk alleles that have been identified in previous studies. We obtained genome-wide methylation profiles and RNA sequencing data for 128 asthmatic and 128 non-asthmatic individuals. We identified DMRs that were enriched for asthma-related genes and were differentially methylated between asthmatics and non-asthmatics. We also identified SNPs that were associated with asthma status in previously published GWAS studies. By integrating these two datasets, we were able to identify a set of 500 DMRs that were significantly associated with asthma status and were also enriched for asthma-related genes. These DMRs were located in promoter regions of genes involved in immune response, inflammation, and epithelial biology. We also identified 200 SNPs that were associated with asthma status and were located in promoter regions of genes involved in immune response, inflammation, and epithelial biology. These results suggest that genetic and epigenetic factors interact in the regulation of gene expression in the airway epithelium of asthmatics.

691F
Investigation of Cytotoxic T-lymphocyte-associated Protein 4 Gene Polymorphisms in Systemic Lupus Erythematosus. H. Yang1, T. Chen2, C. Lee1,2,4, S. Chang3, W. Chen1, W. Lin1, C. Lin1, Y. Lee1,2,6. 1) Dept. of Medical Research, Mackay Memorial Hospital, New Taipei City, Taiwan; 2) Department of Nursing, Mackay Medicine, Nursing and Management College, Taipei, Taiwan; 3) Dept. of Rheumatology, Mackay Memorial Hospital, Taipei, Taiwan; 4) Department of Internal Medicine, Mackay Memorial Hospital, Taipei, Taiwan; 5) Dept. of Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan; 6) Institute of Biomedical Science, Mackay Medical College, New Taipei City, Taiwan; 7) Dept. of Medicine, Mackay Medical College, New Taipei City, Taiwan; 8) Dept. of Pediatrics, Taipei Medical University, Taipei, Taiwan.

Systemic Lupus Erythematosus (SLE) is a chronic inflammatory, multi-systemic, and autoimmune-related disease. The prevalence of SLE in Taiwan is around 37.0/100,000 persons and associated with gender, especially in childbearing age. SLE characterized by T and B cell hyperactivation, autoantibody production, and immune complex deposition. Cytotoxic T-lymphocyte-associated protein 4 (CTLA4) is expressed on CD4+ and CD8+ activated both T and B cells. Protein expression of CTLA4 was increased in T cells from patients with SLE. Numerous studies have been found in association between CTLA4 polymorphisms and SLE in different ethnic groups. However, large-scale study of CTLA4 polymorphisms in SLE is still elusive in Taiwan. For this case-control cross-sectional study among Taiwanese, 283 patients with SLE and 920 controls were enrolled. Genotyping of -318 C/T, +49 A/G, and CT60 A/G single nucleotide polymorphisms (SNPs) was performed by PCR-RFLP and TaqMan assays. The genotype, allele, and haplotype frequencies were calculated by direct counting or with Haploview 4.2 software. Genetic, allele, and haplotype frequencies of the CTLA4 SNPs studied were equally distributed in SLE patients and controls. No significant associations between SLE and these 3 SNPs were observed. Our data suggest that CTLA4 -318 C/T, +49 A/G, and CT60 A/G SNPs do not confer increased susceptibility to SLE.

692W
Large scale identification of alternative polyadenylation through next generation sequencing. C.M. Lopes-Ramos1,2, A.A. Camargo3, P.A.F. Galante4, R.B. Parriguan5. 1) Molecular Oncology, SFPB, Hospital Sírio-Libanês, São Paulo, Brazil; 2) Fundação Antonio Prudente, AC Camargo Cancer Center, São Paulo, Brazil.

Most eukaryotic miRNAs acquire an uncoded polyA tail at their 3' ends during maturation in a process called polyadenylation (polyA). This process involves the cleavage of mRNAs at a specific site and the subsequent addition of adenosine residues. Cleavage position is defined by recognition of specific sequences, called polyA signals. More than half of mammalian genes have multiple polyA signals, which can lead to the formation of transcript variants with different 3'-untranslated regions (UTRs) or even different coding regions. Alternative polyA may influence the location, stability and transport of transcripts, in tissue or disease-specific manner. This is a result of many cis-regulatory elements involved in post-transcriptional regulation located within the 3'UTRs, such as miRNA binding sites. Here we studied alternative polyA events in two colon cancer cell lines (HCT116 and SW480) through next generation sequencing. We developed an original cDNA library protocol to amplify multiple alternative polyadenylation sites in the 3'UTR of mRNAs. We evaluated the expression of the 3'UTRs and identification of polyA sites. Using a SOLID sequencer (Life Technologies), more than 50 million sequences were generated for each cell line. Alternative polyA generating transcripts shorter than the reference transcript (RefSeq with the longest 3'UTR) were verified for more than 6000 genes, corresponding to 30% of genes expressed in these cells. Besides the above-mentioned effects, evidences of alternative polyA functional impact may be demonstrated by the loss of miRNA target sites. In fact, most miRNAs have at least one conserved miRNA target site, which represent about 30% of all identified variants. Preferential usage of shorter polyA variants allows escaping from the inhibitory regulation of miRNAs on the expression of such target genes. We have also evaluated miRNA expression profile of these two cell lines and identified more than 50 miRNAs differently expressed in polyA expression profile, suggesting that polyA variants expression might be tightly associated. Therefore, identification of inverse correlation of miRNA and polyA variants expression is underway. This innovative approach increased our capacity of identification and expression of alternative polyadenylation variants. Large scale studies of alternative polyadenylation sites in other cancers might provide new tools to understand how this phenomenon regulates different cellular processes in physiological and pathological situations.
693T
Identification of DNA editing in retrotransposons of diverse genomes.
B.A. Knisbacher, E.Y. Levanon. The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel.

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

694F
Association between estrogen receptor and intraocular pressure in mice. F. Mabuchi1, R. Yamagishi2, K. Kashiwagi1, M. Aihara3. 1) Dept Ophthalmology, Univ Yamanashi, Chuo, Yamanashi, Japan; 2) Dept Ophthalmology, Univ Tokyo, Bunkyo-ku, Tokyo, Japan; 3) Shirato eye clinic, Shinjuku-ku, Tokyo, Japan.

Purpose: To evaluate intraocular pressure (IOP) in transgenic mice with a knockout of estrogen receptor alpha (ESR1) or estrogen receptor beta (ESR2) gene. Methods: Heterozygous B6.129P2-Esr1tm1Ksk/J (ESR1 gene knockout) or B6.129P2-Esr2tm1Unc/J (ESR2 gene knockout) mice, and corresponding wild-type mice were anesthetized. A fluid-filled glass microneedle connected to a pressure transducer was then inserted through the cornea into the anterior chamber to measure IOP. All measurements were made between 9:00 PM and 11:00 PM. The IOP of fifteen ESR1+/- and wild-type ESR1+/+ mice, and IOP of nine ESR2+/- and wild-type ESR2+/+ mice were measured at 6 to 8 weeks after birth. Results: The IOPs of the ESR1+/- and wild-type ESR1+/+ mice were 20.5 ± 1.0 mmHg (mean ± standard deviation) and 20.3 ± 1.2 mmHg respectively, and there was no statistically significant difference (P = 0.81, Mann-Whitney U test) between them. In contrast, the IOPs of the ESR2+/- and wild-type ESR2+/+ mice were 18.3 ± 1.2 and 19.9 ± 1.6 mmHg respectively, and there was a statistically significant difference (P = 0.04, Mann-Whitney U test) between them. Conclusion: These results suggest that the ESR2 is associated with IOP regulation.
The role of genome copy number variation of the CCL3L1 gene, encoding MIP1α, in contributing to the host variation in susceptibility and response to HIV infection is controversial. Here we analyse a sub-Saharan African cohort from Tanzania and Ethiopia, two countries with a high prevalence of HIV-1 and a high co-morbidity of HIV with tuberculosis. We use a form of quantitative PCR called the paralogue ratio test, to determine CCL3L1 gene copy number in 1134 individuals and validate our copy number typing using array comparative genomic hybridisation and fish-FISH. We find no significant association of CCL3L1 gene copy number with HIV load in antiretroviral-naïve patients prior to initiation of combination highly active anti-retroviral therapy. However, we find a significant association of low CCL3L1 gene copy number with improved immune reconstitution following initiation of highly active anti-retroviral therapy (p=0.012), replicating a previous study. Our work supports a role for CCL3L1 copy number in immune reconstitution following antiretroviral therapy in HIV, and suggests that the MIP1α-CCR5 axis might be targeted to aid immune reconstitution.
698W
Differential effects at candidate SNP loci on reduction in plasma LDL cholesterol, apolipoprotein B, LDL particle number, and mean LDL particle size with statin therapy. J. F. Chu1, B. L. Barratt2, B. Ding3, F. Nyberg1,4, S. Mora1,5, P. M. Ridker5, D. I. Chasman1,6. 1) Division of Preventive Medicine, Brigham & Women’s Hospital, Boston, MA USA; 2) Personalized Healthcare and Biomarkers, AstraZeneca Research and Development, Alderley Park, United Kingdom; 3) Global Epidemiology, AstraZeneca Research and Development, Molndal, Sweden; 4) Unit of Occupational and Environmental Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 5) Division of Cardiology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA USA; 6) Division of Genetics, Brigham and Women’s Hosp, Boston, MA USA.

Statins differentially reduce LDL cholesterol (LDL-C), apolipoprotein B (apoB), total number of LDL particles (LDL-P) and mean LDL particle size (LZ). Variations in the reduction of these measures and possibly their relationship to cardiovascular risk reduction may be under genetic control. We performed a candidate SNP analysis for association with change in LDL-C, apoB, LDL-P and LZ after 12 months of statin therapy, focusing on 156 SNPs previously associated with either statin-induced reduction in LDL-C (6 SNPs) or untreated serum lipid levels from meta-analysis by the Global Lipids Consortium (150 SNPs). The study population included 3,534 statin-associated participants of European ancestry in JUPITER (NCT00239681), a randomized, placebo-controlled trial of rosvustatin (20 mg/d) for primary prevention of cardiovascular disease. Residuals for all LDL change measures were calculated adjusted for age, sex, region and population stratification and further transformed by inverse-variate normalization. Genetic associations were assessed by linear regression applying age and population stratification as covariates. Assuming MAF of 0.15, we had more than 80% power to detect a mean difference of 2.4% between carriers and non-carriers of the risk genotype in the additive model. Several multi-SNP analyses were also carried out, including haplotype analysis, stepwise linear regression, and gene-gene interactions. The statistical strength of our biological hypothesis was measured by comparing the sum of the observed association evidence across all 60 SNPs with the value expected under the null based on a phenotype permutation method (10,000 permutation replicates). We observed significant associations between rs8435326 in NDUFS1 and percentage of weight gain (P = 0.0037) even after adjusting for age, gender, and body mass index (BMI) at baseline. Additionally, we observed significant associations between the TT risk genotype of rs8435326 in NDUFS1 and AG genotype of rs3762883 in COX18 (Pcorr=0.001). Finally, permutation-based test showed that the set of 60 SNPs from the 28 nuclear-encoded mitochondrial genes selected based on our hypothesis, collectively, was associated with weight gain (% (P = 0.005). In addition, we demonstrated higher cumulative attainment rates of target fasting plasma glucose levels (Plog-rank = 0.0009) than the A allele carriers, and this difference was still significant (P = 0.0037) even after adjusting for age, gender, and body mass index (BMI) at baseline. Besides, significant linear relationships were detected between the number of rs10229583 A alleles and increments in fasting insulin levels (P= 0.0426, adjusted for age, gender, dosage, and BMI at baseline). However, no effect of rs10229583 on the efficacy of rosvustatin was found in our study. In conclusion, the PAX4 variant rs10229583 was associated with therapeutic effect of repaglinide in Chinese patients with type 2 diabetes.

700W
PAX4 genetic variant is associated with therapeutic effect of repaglinide in the Chinese type 2 diabetes patients. C. Hu, M. Chen, R. Zhang, Y. Bao, K. Xiang, W. Jia. Shanghai Diabetes Inst, Shanghai Jiao Tong University, Shanghai, China.

A recent genome-wide association study (GWAS) identified a novel diabetes-associated locus near PAX4 (rs10229583) in Chinese population and other populations. However, whether this single-nucleotide polymorphism (SNP) influence the therapeutic effects of oral antihyperglycemic drugs has not been reported. The aim of this study was to investigate the association of PAX4 variants rs10229583 with therapeutic effects of repaglinide or rosiglitazone in patients with type 2 diabetes. A total of 209 newly diagnosed type 2 diabetes patients were recruited and treated with repaglinide or rosiglitazone randomly for 48 weeks (104 and 105 patients, respectively). In the repaglinide cohort, individuals who were GG homozygotes of rs10229583 showed lower fasting plasma glucose, 2h glucose and HbA1c levels than the A allele carriers. Moreover, carriers of GG genotype exhibited significantly higher cumulative attainment rates of target fasting plasma levels (Plog-rank = 0.0009) than the A allele carriers, and this difference was still significant (P = 0.0037) even after adjusting for age, gender, and body mass index (BMI) at baseline. Besides, significant linear relationships were detected between the number of rs10229583 A alleles and increments in fasting insulin levels (P= 0.0426, adjusted for age, gender, dosage, and BMI at baseline). However, no effect of rs10229583 on the efficacy of rosiglitazone was found in our study. In conclusion, the PAX4 variant rs10229583 was associated with therapeutic effect of repaglinide in Chinese patients with type 2 diabetes.
Multi-ethnic cytochrome P450-2D6 (CYP2D6) allele frequency profiling and full gene single-molecule real-time (SMRT®) sequencing. S. Mantis1, Y. Yang1, A. Gaedigk2, R. Sebra3, R. Vizelaar4, R. Kornreich1, R.J. Dennis5, S.A. Scott1. 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Division of Clinical Pharmacology & Therapeutic Innovation, University of Missouri-Kansas City, Kansas City, MO; 3) MRC Holland, Willem Schoutenstraat 6, Amsterdam, The Netherlands.

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

Association of Cyclooxygenase-2 genetic variant with cardiovascular disease. S.A. Ross1, J.W. Eikelboom2, S.S. Anand1, N. Eriksson2, H. Gerstein1,4, S.R. Mehta1, S.J. Connolly1, L. Rose6, P.M. Ridker5,6, L. Wallentin2, D. Chasman5,6, S. Yusuf1, G. Pare1. 1) Population Health Research Institute, Hamilton Health Sciences and Departments of Medicine, Epidemiology, Pathology McMaster University, Hamilton, ON, Canada; 2) Uppsala Clinical Research Center and Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 3) Department of Medicine, McMaster University, Hamilton, ON, Canada; 4) Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada; 5) Division of Preventive Medicine, Brigham and Women’s Hospital, Boston, Massachusetts, USA; 6) Harvard Medical School, Boston, Massachusetts, USA.

Background: Cyclooxygenase (COX) enzymes convert arachidonic acid to prostaglandin H2 leading to the production of biologically active prostanoids, prostacyclin and thromboxane, which are believed to play key roles in atherothrombosis. The non-selective COX inhibitor aspirin is associated with a decreased risk of cardiovascular disease (CVD) whereas selective COX-2 inhibitors have been associated with an increased risk. A genetic variant (rs20417) of the PTGS2 gene, encoding for COX-2, has been associated with increased COX-2 activity and a decreased CVD risk. However, this genetic association and the role of COX-2 in CVD remains controversial.

Methods: The association between rs20417 and major adverse cardiovascular outcomes (non-fatal myocardial infarction, stroke or cardiovascular death) was prospectively explored in 49,233 subjects (ACTIVE-A, CURE, epiDREAM/DREAM, ONTARGET, RE-LY, and WGS) and the potential of modifiable risk factors on the genetic association was further explored in 4,465 non-fatal myocardial infarction cases and 4,898 controls from the INTERHEART study. We also examined the effects of rs20417 on urinary thromboxane and prostacyclin metabolite concentrations in 119 healthy individuals. Results: Carriage of the rs20417 minor allele was associated with a decreased risk of major CVD outcomes (OR=0.78, 95% CI: 0.70 - 0.87; P=1.2×10−5). The genetic effect was significantly stronger in aspirin users (OR: 0.74, 95% CI: 0.64-0.84; P=1.20×10−5) vs. non-users (OR: 0.87, 95% CI: 0.72-0.84; P=0.16, N=22,441) with an interaction p-value of 0.0041. Carriers had significantly lower urinary levels of thromboxane (97.0 vs. 125.5 ng/mmol creatinine; P=0.02) and prostacyclin (3336.0 vs. 4702.0 ng/mmol creatinine; P=0.01) metabolites as compared to noncarriers. Conclusion: The rs20417 polymorphism is associated with a reduced risk of major cardiovascular events and lower levels of thromboxane and prostacyclin. Our results suggest that a genetic decrease in COX-2 activity may be beneficial with respect to CVD risk and may vary in effect size depending on ASA use.
703W Influence of CYP4F2 on dose, anticoagulation control and risk of hemorrhage among African American and European American warfarin users. A. Shendre1, T.M. Beasley2, D.A. Nickerson3, N.A. Lindsly4, 1. Epidemiology, University of Alabama at Birmingham, Birmingham, AL; 2) Biostatistics, Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, AL; 3) Genome Sciences, School of Medicine, University of Washington, Seattle, WA; 4) Neurology, University of Alabama at Birmingham, Birmingham, AL.

The management of warfarin therapy is complicated by a wide variation in dose and response across patients. A significant portion of variability in dose is explained by single nucleotide polymorphisms (SNPs) in CYP2C9 and VKORC1 across racial groups. CYP4F2 (rs21086622:p.V433M) explains variability in dose among European Americans but its influence on warfarin dose, anticoagulation control, and risk of hemorrhage has yet to be fully examined, especially in African Americans. Herein we assess the influence of CYP4F2 on stable warfarin dose, anticoagulation control (percent time in target range; PTTR) and risk of major hemorrhage in 649 European and 430 African Americans patients in a prospective warfarin pharmacogenetics study. We evaluated the influence of the CYP4F2 genotypes on the 3 outcomes of interest using an additive model with adjustment for clinical factors, co-medications, and established genetic predictors (i.e., VKORC1; CYP2C9). For warfarin dose and anticoagulation control (PTTR) as outcomes, multivariable linear regression modeling was performed. The association of CYP4F2 with risk of hemorrhage was examined using Cox proportional hazards regression. The frequency of the minor (A) allele was higher in European (52%) compared to African (16%) Americans. Compared to the GG genotype, the CYP4F2 A allele was associated with a 7.0% higher (p=0.03) hemorrhage risk. The CYP4F2 C allele was associated with a 6.9% higher (p=0.03) hemorrhage risk compared to the AA genotype. These data further contribute to the literature on clinical decision making and to establish the utility of CYP4F2 in the care of a racially diverse patient population.

704W A mechanism for docetaxel induced neutropenia: The role of Cwc27. T. Wiltshire1, C.S. Benton2, F. Muhale3, A. Frick3, O.J. Trask3, R. Thomas2, S. Cai4, H.L. McLeod1, 1) School of Pharmacy, University of North Carolina, Chapel Hill, NC; 2) The Hamner Institutes for Health Sciences, Research Triangle Park, NC; 3) Washington University School of Medicine, St. Louis, MO.

Chemotherapy-induced neutropenia (CIN) often results in dose reduction or treatment delay, negatively impacting response rates and overall survival for many cancer patients. Although several studies have suggested that genetic factors play a role in CIN, and genetic information could be used to better identify patients who are likely to develop CIN prior to drug and dose selection. To identify genes that mediate docetaxel-induced neutropenia, we conducted a genome-wide association analysis in twenty-one genetically diverse mouse inbred strains. After a 14-day treatment with docetaxel, we found a significant association warfarin dose in African Americans (0.1% higher (p=0.98) for AG and 13.6% higher (p=0.6) for AA genotype). PTTR was not influenced by CYP4F2 for either European American (p=0.59) or African American patients (p=0.78). The incidence of hemorrhage was lowest among patients with the AA genotype (1.05/100yrs) compared to the AG (8.1/100yrs) and GG (6.8/100yrs) genotype. Patients with AA genotype had a lower risk of hemorrhage (HR: 0.52-0.37; p=0.09) compared to patients with AG or AG genotype. Incorporation of CYP4F2 genotype can provide additional improvement in warfarin dosing among European Americans. Additionally, CYP4F2 AA genotype appears to provide a protective effect against the risk of hemorrhage further contributing to the literature on clinical decision making and to establish the utility of CYP4F2 in the care of a racially diverse patient population.

705W ABCB1 rs3842 polymorphism affects tacrolimus pharmacokinetics during the first week after liver transplantation in Chinese. J. Shi1, C. Zhang1, D. Chen2, K. Zhang3, Z. Peng4, W. Zhang5, 1) Department of Genetics, Chinese National Human Genome Center at Shanghai, Shanghai, China; 2) Department of General Surgery, Shanghai First People’s Hospital, Shanghai Jiao Tong University, Shanghai, China; 3) Rui Jin Hosp. School of Medicine, Shanghai Jiao Tong University, Shanghai, China.

Immunosuppressive drug, tacrolimus, is widely used in solid organ transplantation to prevent allograft rejection. However, it has a narrow therapeutic index with high inter-individual variations in its pharmacokinetics, which makes this drug a potential candidate for pharmacogenetic research. We have studied the correlation between the ABCB1 genotype polymorphisms and tacrolimus pharmacokinetic. Moreover, the obtained results are conflicting. The reason for these conflicting results might be the fact that the studies were mostly focused on 12436C>T (rs1128503), 2677G>T/A (rs2032582) and 3435C>T (rs1045682). However, these polymorphisms are not the only pharmacokinetics capable of influencing the P-glycoprotein function. Thus, we selected ten variations with minor allele frequency greater than 3% in Han Chinese and examined genotypes for both recipient and donor in 115 liver transplantation patients, to clarify the influence of these genetic variants on tacrolimus requirements. The recipient ABCB1 variant rs3842 significantly influenced the mean tacrolimus trough blood concentrations (P=2.6×10⁻³), and the mean dose-adjusted trough concentrations (P=8.8×10⁻³) during the first week post transplantation. A significantly higher trough blood concentration was observed for recipients with the homozygous AA genotype (−logP >3.0) compared to the homozygous AA genotype. This association remained significant after adjusting for age, sex, clinical outcome parameters and information of combined drugs. None of the donor ABCB1 polymorphism showed correlation with tacrolimus pharmacokinetics. The extensively studied 3435C>T, 2677G>T/A and 12436C>T in both donor and recipient showed no correlation with tacrolimus pharmacokinetic. This study indicates, for the first time, that recipient ABCB1 polymorphism rs3842 is associated with tacrolimus dose required during the early post-transplant period. The result highlights the importance of screening the comprehensive polymorphisms in ABCB1 gene region and examining the impact of polymorphisms in different post transplant time, which might provide new clues of the association of ABCB1 and tacrolimus pharmacokinetic.
707W

The Role of Multidrug Resistance-1 (MDR1) Variants in Response to Atorvastatin among Jordanians. K.H. Alzoubi1, O.F. Khabour1, S.A. Alazzam1. 1) Clinical Pharmacy, Jordan University of Science and Technology, Irbid, Jordan; 2) Medical Laboratory Sciences, Jordan University of Science and Technology, Irbid, Jordan.

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

708W

Association of ABCB1, SLC15A1, CES1 and NEU2 gene polymorphisms with side effects by oseltamivir in a Mexican population. M. Bermudez de Leon 1, 2, R.N. Gonzalez-Rios1, 2, R.A. Leyva-Parr³, E. Gonzalez-Gonzalez1, 2, A. Alvarado-Diaz1, 2, O.E. Vazquez-Monsivais2, V.L. Mata-Tijerina2, B.L. Escobedo-Guajardo2, M.J. Currás-Tula2, M.E. Aguado-Barrera2, B. Silva-Ramirez2, J.M. Alcocer-Gonzalez1, A.M. Salinas-Martinez2. 1) Dept of Mol Biol, Centro de Investigación Biomédica del Noreste. Instituto Mexicano del Seguro Social, 64720 Monterrey, Nuevo León, Mexico; 2) Laboratory of Molecular Diagnostics, Centro de Investigación Biomédica del Noreste. Instituto Mexicano del Seguro Social, 64720 Monterrey, Nuevo León, Mexico; 3) Dept of Immunogenetics, Centro de Investigación Biomédica del Noreste. Instituto Mexicano del Seguro Social, 64720 Monterrey, Nuevo León, Mexico; 4) School of Biological Sciences, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Nuevo León, Mexico; 5) Epidemiology Research and Health Services Unit, Instituto Mexicano del Seguro Social, 64000 Monterrey, Nuevo León, Mexico; 6) School of Pharmacy, University of California, Berkeley, CA; 7) University of California, San Francisco, CA; 8) Baylor College of Medicine, Houston, TX; 9) Children's Hospital Oakland Research Institute, Oakland, CA; 10) Ohio State University, Columbus, OH; 11) Harvard School of Public Health, Boston, MA.

Oseltamivir is a neuraminidase inhibitor extensively used during the last pandemic influenza to prevent the release of progeny virions and thereby limit the spread of infection. Oseltamivir phosphate is a prodrug that is absorbed by Peptide transporter PEPT1 and effluxed by P-glycoprotein. Once inside of the cell, oseltamivir phosphate is converted by human carboxylesterase 1 to oseltamivir carboxylate, which is the active form to inhibit influenza virus neuraminidase. However, oseltamivir also has an inhibitory effect on human sialidases, that are important in various cellular functions including lysosomal catabolism. There are reports about deaths and neuropsychiatric events in Japanese population with the use of oseltamivir, suggesting that this drug could inhibit also to human sialidases in a similar way that viral sialidase. Thus, the aim of this study was to determine the single nucleotide polymorphisms (SNPs) for PEPT1 (SLC15A1 gene), P-glycoprotein (ABCB1 gene), carboxylesterase 1 (CES1 gene) and sialidase (NEU2 gene) in >700 Mexican patients with oseltamivir therapy, and these data were correlated with side effects reported between 2010-2012. The SNPs evaluated were Gly185Val in SLC15A1 (p = 0.02), Gly143Glu and Arg199His in CES1 (p < 0.001) and 2907243, respectively for CES1 gene; and Arg41Gln in NEU2 gene. Clinical data of each patient were obtained from institutional electronic files. This work was approved by the Ethics Committee of Mexican Social Security Institute. We found that eight percent of patients showed side effect as depression, anxiety, seizures, hallucinations among others. Genetic and allele frequencies are presented and associated with the side effects of oseltamivir in a Mexican population. The authors thank Consejo Nacional de Ciencia y Tecnologia for financial support (Grant number SALUD-2011-1-162243).

709W

PGRN Network-wide Project: Transcriptome Analysis of Pharmacogenes in Human Tissues. E.R. Gamazon1, C.E. French2, S.W. Yee3, A. Chinnbri1, X. Qin4, E. Theusch3, A. Webb4, A. Konkashbaev1, S. Weiss2, S.B. Brenner2, S.E. Scherer1, N.J. Cox1, K.M. Giacomini5. 1) University of Chicago, Chicago, IL; 2) University of California, Berkeley, CA; 3) University of California, San Francisco, CA; 4) Baylor College of Medicine, Houston, TX; 5) Children's Hospital Oakland Research Institute, Oakland, CA; 6) Ohio State University, Columbus, OH; 7) Harvard School of Public Health, Boston, MA.

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

710W

Systematical functional characterization of CYP2D6 alleles in the Chinese Han population. S. Jin1, O. Xu2, Z. Wu2, L. He3, Bio-X Institutes, Shanghai Jiaotong University, Shanghai, Shanghai, China.

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

1) Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI; 2) Department of Public Health Sciences, Henry Ford Health System, Detroit, MI; 3) Department of Medicine, University of California, San Francisco, CA; 4) Department of Bioengineering & Therapeutic Sciences, University of California, San Francisco, CA; 5) Department of Pediatrics, The Ann and Robert H. Lurie Children’s Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, Illinois, Chicago, University of California, San Francisco, CA; 7) Cardiovascular Research Institute, University of California, San Francisco, CA; 8) Department of Medicine, Henry Ford Health System, Detroit, MI.

Inhaled short-acting beta2-agonists (SABAs) are commonly used in patients with asthma to rapidly reverse airway obstruction and improve acute symptoms. However, treatment response to SABA medication is highly variable and is likely to be genetically influenced. We performed a genome-wide association study of SABA medication response in 326 healthy, non-asthmatic African Americans and replicated these findings in 1073 individuals with asthma and 149 without asthma. Healthy, non-asthmatic individuals were used for discovery as SABA response might be less obscured by unrelated factors associated with lung disease. A linear mixed model approach was first used for SNP associations, and results were then combined to generate gene-based associations. We also assessed whether genes associated with SABA response were related to the amount of SABA use among those with asthma. Gene-based association tests identified SPATA13-AS1 on the X chromosome as an antisense RNA encoding segment within the RNA gene as being significantly associated with SABA response. In replication, this gene was also associated with SABA response among 1073 African American individuals with asthma (p = 0.011). The same association was also replicated in an additional 149 healthy African American individuals (p = 0.027). Lastly, SPATA13-AS1 was also associated with annual SABA medication use among individuals with asthma (p = 0.047). SPATA13 is a recently discovered protein that is likely to be involved in airway smooth muscle contraction-relaxation through its interactions with Rho family GTPases such as RhoA, Rac1 and Cdc42. Elucidating the precise mechanism of action of this protein may reveal new SABA response pathways as well as targets for future asthma therapeutics.

712W

Genome-wide analysis of Methotrexate pharmacogenomics in rheumatoid arthritis reveals novel risk variants and leads for TYMS regulation.


1) Dept Genetics, Univ Delhi, South Campus, New Delhi, Delhi, India; 2) National Institute of Immunology, New Delhi, India; 3) Department of Rheumatology, Fortis Flt. Lt. Rajan Dhall Hospital, New Delhi, India; 4) Division of Rheumatology & Clinical Immunology, Medanta Bone & Joint Institute, Medanta - The Medicity, Gurgaon, India; 5) Department of Rheumatology, All India Institute of Medical Sciences, New Delhi, India; 6) Department of Physics, University of Delhi, Delhi, India.

Objective: Methotrexate (MTX) is the drug of first choice for rheumatoid arthritis as an antirheumatic but is effective only in around 60% of the patients. Further, early initiation of MTX in a brief critical window period of around 6 months is crucial for the effective treatment. Therefore, identification of predictive markers is critical for the choice of early and appropriate treatment regimen. To date, genetic determinants underlying MTX response have remained elusive despite extensive candidate gene association studies. In the present study, we used genome-wide genotype data to identify potential risk variants associated with MTX (non)response in a north Indian RA cohort.

Methods: Genome-wide genotyping data from Illumina 660w quad array for a total of 457 RA patients [297 good (DAS 28-3 < 3.2) and 160 poor (DAS 28-3 >5.1)] responders on MTX monotherapy were tested for association using additive model. Other tools including support vector machine (SVM) and genome-wide pathway analysis were used to identify additional risk variants and pathways. All risk loci were imputed to fine-map the association signals and identify causal variant(s) of therapeutic/diagnostic relevance.

Results: Nine novel suggestive loci from GWAS (p≤x10−5) and two from SVM analysis were associated with MTX (non)response. Association of published genome-wide SNPs with MTX response was reaffirmed. Imputation followed by bioinformatic analysis indicated possible interaction between two reversely oriented overlapping genes namely ENOSF1 and TYMS at post-transcriptional level. Conclusion: In this first ever genome-wide analysis on MTX treatment response in RA patients, 11 new risk loci were identified. Further, TYMS expression at post-transcriptional level seems probably regulated through an antisense-RNA involving the Bop insertion marker in the overlapping segment at 3′ UTR of TYMS, ENOSF1, a finding with impending pharmacogenetic applications.

713W

Using Neuroimaging Endophenotypes to Identify Molecular Markers for Treatment Response to Major Depressive Disorder.


1) Max Planck Institute for Psychiatry, Munich, Germany; 2) Department of Mathematics, University of Tulsa; 3) Department of Psychiatry and Behavioral Sciences, School of Medicine, Emory University; 4) Department of Biostatistics and Bioinformatics, School of Public Health, Emory University; 5) Department of Psychiatry and Behavioral Sciences and Department of Neurology, School of Medicine, Emory University.

Major depressive disorder (MDD) is a prevalent disease with high rates of treatment resistant and non-remission. However biological measures to guide optimal treatment have been lacking. Recently, our group described that resting state brain activity patterns (BAPs) of specific brain regions (i.e. right insula, right inferior temporal cortex, left amygdala, left premotor cortex, right motor cortex and precuneus) can predict differential response to either escitalopram (sCIT) or psychotherapy. The aim of this study was to identify molecular markers that associate with these BAPs, in the hope to identify predictive measures that are more easily obtained in clinical practice than neuroimaging measures. Patients were recruited at Emory University and randomized at baseline to either 12 weeks sCIT, or 16 sessions of cognitive behavioral therapy (CBT). Pre-treatment BAPs of the six brain regions of interest (ROIs) were used as neuroimaging endophenotypes for this study. In peripheral blood DNA drawn at baseline genome-wide SNP genotypes (Illumina HM 450K) were measured. Genome-wide univariate and multivariate association analyses including all possible ROIs combinations were conducted in 76 MDD patients. Tests for association between methylation status at ~485,000 CpG sites and BAPs was observed genome-wide significant association of rs34383296, (p = 9.4x10−9) in a multivariate analysis that included the right insula, left amygdala and left prefrontal cortex. Univariate analyses did not reveal genome-wide significant associations. Further, single nucleotide variation in a gene dense region on chromosome 9 within the NDOR1 gene locus and it is an eQTL for ARRDC1, a gene ~400kb downstream and is related to alcohol-mediated internalization of cell surface receptors. No genome-wide significant association of DNA-methylation status with BAPs was observed. The best candidate for further processing is a gene with putative significant association with neuroimaging endophenotypes and genomic approaches may be able to identify markers to guide individualized depression therapy choices in clinical routine. Further analysis will test surrounding CpG sites for the candidate gene as well as mQTLs as well as imputation of additional variants in the locus and tests for association with treatment response in independent samples.
714W

Genome-wide association study loci are enriched for clinically relevant drugs targets for common human diseases. R. Li1, V. Forgetta2, O. Yu2, Z. Dastani1, M. Lathrop3, J.B. Richards4,5. 1) Departments of Medicine, McGill University, Montreal, Quebec, Canada; 2) Departments of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada; 3) Department of Medicine, McGill University, Montreal, Quebec, Canada; 4) Departments of Epidemiology, Biostatistics and Occupational Health, Jewish General Hospital, Lady Davis Institute, McGill University Montreal, Quebec, Canada; 5) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 6) Twin Research and Genetic Epidemiology, King’s College London, London, United Kingdom.

While many genetic loci have been recently mapped to common human disease it is not yet apparent if this information will aid in drug development. We reasoned that if GWAS loci identified such targets then these loci would be enriched for the targets of drugs that are already on the market. To test this hypothesis, we determined enrichment of GWAS loci for clinically relevant drugs for multiple common human diseases. We selected easily definable, representative, common diseases, with relatively high heritabilities for which large-scale GWAS meta-analyses had been published. These included, type 2 diabetes, hypercholesterolemia (LDL), osteoporosis, Crohn’s disease, hypertension, psoriasis and asthma. We also included an additional disease, COPD, whose current therapies are largely aimed at symptom control. Pipeline was used to identify marketed drugs and their targets and this list was refined to include only drugs accepted by relevant American clinical guidelines. The NHGRI GWAS Catalog was used to collect all genome-wide significant SNPs for each disease and genes within 100kb or 500kb were identified. The nearest network neighbours for each drug target were identified using STRING. Finally, since several immune-mediated diseases have promising interleukin-based therapies in development we assessed whether drugs in phase III of development were enriched at GWAS loci. On average, there were 9 drug targets per disease that achieved guideline-level acceptance. Assessing all genes within 500kb of GWAS SNPs and the nearest network neighbours of drug targets, 80% of guideline-accepted LDL drug targets were identified. The proportion of such targets identified for osteoporosis was 90%, 71% for type 2 diabetes, 43% for Crohn’s, 22% for psoriasis, 15% for hypertension, 9% for asthma and 0% for COPD. Consideration of Phase III drugs improved these proportions to 56% for Crohn’s, 45% for psoriasis and 29% for asthma. Genes within 100kb of GWAS SNPs were strongly enriched for guideline-accepted drug targets for some diseases (eg. 40-fold [95% CI: 14-118] for LDL and 35-fold [95%CI: 12-104] for osteoporosis). GWAS loci for some common diseases identify a large proportion of drug targets that are targeted by clinically relevant medicines. This implies that other clinically relevant drug targets exist among the novel loci identified from GWAS.

715W

Conditional analysis using HLA-A*31:01 as a covariant to detect additional genetic risk factors for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. T. Ozeki1, T. Mushiroda1, A. Takahashi2, M. Kubo3, Y. Shirakata4, Z. Ikezawa5, M. Iijima6, T. Shiohara7, K. Hashimoto8, Y. Nakamura9. 1) Laboratory for Pharmacogenomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan; 2) Laboratory for Stastical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan; 3) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan; 4) Department of Dermatology, Ehime University Graduate School of Medicine, Ehime, Japan; 5) Department of Dermatology, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 6) Department of Dermatology, Showa University School of Medicine, Tokyo, Japan; 7) Department of Dermatology, Koryo University School of Medicine, Tokyo, Japan.

An anticonvulsant, carbamazepine (CBZ), is known to show incidences of cutaneous adverse drug reactions (cADRs) including Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) and drug-induced hypersensitivity syndrome (DIHS). Previously, to identify a gene(s) susceptible to CBZ-induced cADRs, we conducted a genome-wide association study (GWAS) and following HLA typing in 61 cases and 376 patients who showed no cADRs by administration of CBZ (CBZ-tolerant controls), and found that HLA-A*31:01 was present in 60.7% (37/61) of the patients with CBZ-induced cADRs, but in only 12.5% (47/376) of the CBZ-tolerant controls (odds ratio (OR) = 10.8, 95% confidence interval (CI) = 5.9–19.6, \(P = 3.64 \times 10^{-15}\)). To confirm associations of HLA-A alleles other than HLA-A*31:01 with CBZ-induced cADRs, we pick-upped a paired HLA-A allele of HLA-A*31:01 using case-control subjects carrying at least one HLA-A*31:01 allele, which consisted of 56 CBZ-induced cADR cases and 54 CBZ-tolerant controls. There was no HLA-A allele that reached significant association after Bonferroni correction with \(P < 4.17 \times 10^{-3}\) (0.05/12 HLA-A alleles). To detect additional genetic marker for CBZ-induced cADRs, we performed genome-wide imputation using the GWAS data and genotype data from 1000 Genomes Project as reference panels by MaCh-Admix software. For a conditional logistic regression analysis of the imputed genotype data, we selected rs1633021 as the covariant with the firmest association observed in the analyzed GWAS dataset. After the conditional analysis, there were 22 loci that reached GWAS-level significant association \(P < 5 \times 10^{-8}\). The replication analysis using independent population are ongoing.
716W
Whole genome association study identifies novel antidepressant response loci for the treatment of obsessive-compulsive disorder with selective serotonin re-uptake inhibitors. H.D. Qin2, Y. Wang2, M.A. Grados2, M.A. Riddle2, B.D. Greenberg4, J.A. Knowles5, A.J. Fyer6, J.T. McCracken7, D.L. Murphy8, S.A. Rassmusen4, B. Cullen9, J. Piacentini2, D. Geller2, D. Pauls2, E. Stewart10, O.J. Bienvenu11, Y. Chen9, F.S. Goess9, B. Maher12, J.P. Samuels2, G. Nestadt2, Y.Y. Shugart1. 1) Unit of Statistical Genomics, Intramural Research Program, Division of Intramural Research Program, National Institute of Mental Health, NIH, Bethesda, MD 20892, USA; 2) Department of Psychiatry and Behavioral Sciences, School of Medicine, Johns Hopkins University, Baltimore, Maryland; 3) Departments of Psychiatry and Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA; 4) Department of Psychiatry and Human Behavior, Brown Medical School, Butler Hospital; 5) Department of Psychiatry, Keck Medical School, University of Southern California; 6) College of Physicians and Surgeons at Columbia University; 7) Department of Psychiatry and Behavioral Sciences, School of Medicine, University of California, Los Angeles; 8) Laboratory of Clinical Science, NIMH, NIH, Bethesda; 9) Department of Medicine, University of Maryland School of Medicine; 10) Department of Psychiatry, University of British Columbia, A3-118, West 28th Avenue, Vancouver, BC, Canada V5Z 4H4. Selective serotonin reuptake inhibitors (SSRIs) are first line medications for the treatment of obsessive-compulsive disorder (OCD). Although SSRIs are currently the most frequently used drug therapy for OCD, approximately 30% of OCD patients show limited or no response to these medications, and >77% cannot tolerate side effects. Genetic predictors for OCD treatment response have been identified through the analysis of a large number of common variants with small effects. The top two significant variants are located in six genes, namely DLGAP1 and CSMD1, which may play an important role in SSRI drug response in OCD patients. The importance of identification of drug response loci is for the development of ‘personalized medical treatment’ of OCD patients treated with SSRIs. The potential results would provide also new targets for developing novel drugs for the treatment of non-responders. Further well-designed case-control studies with large sample size and using next-generation sequencing are needed to explore the role of causal exonic variations or rare CNVs of the most significant genes.

717W
Integrative genome modeling reveals common genetic architecture of neuropathy resulting from distinct environmental exposures. H.E. Wheeler1, C. Wing1, M. Komatsu1, S. Delaney1, E.R. Gamazon2, C. Rodriguez-Antona3, N.J. Cox4, M.E. Dolan5. 1) Dept Medicine, University of Chicago, Chicago, IL; 2) Human Cancer Genetics Programme, Spanish National Cancer Research Center, Madrid, Spain. Chemotherapy-induced peripheral neuropathy is the major dose-limiting toxicity for several mechanistically distinct anticancer drugs. Diabetic neuropathy affects a large proportion of diabetes patients due to prolonged exposure to high levels of glucose. Our goal is to elucidate the molecular genetic mechanisms underlying such neuropathies by integrating results from patient and lymphoblastoid cell line (LCL) genome-wide association studies (GWAS) with functional results from iCell Neurons derived from human induced pluripotent stem cells (iPSCs). When comparing modestly sized pharmacogenomic GWAS from patients and LCLs treated with the same drug, SNPs rarely overlap at stringent thresholds such as \( P < 10^{-6} \), but significant overlaps of SNPs at more relaxed thresholds determined by enrichment analysis through random sampling are possible. Under this cumulative hypothesis, large numbers of common variants with small effects account for substantial heritability. For example, we observed an enrichment of paclitaxel-induced LCL (\( n = 247 \)) cytotoxicity SNPs in the peripheral sensory neuropathy-associated SNPs from ovarian and lung cancer patients (\( n = 143 \)) treated with paclitaxel and carboplatin (empirical \( P = 0.034 \)). Interestingly, we also observed an enrichment of paclitaxel-cytotoxicity SNPs in diabetic neuropathy-associated SNPs from patients (\( n = 1651 \)) in the GoKinD cohort (empirical \( P < 0.001 \)). These enrichments demonstrate that susceptibilities to increased cytotoxicity in LCLs and increased chemother-apy- and diabetes-induced neuropathy in patients likely have some genetic mechanisms in common. We are using iPSC-derived iCell Neurons to develop models for functional screens of candidate targets of interest from these enrichment analyses. Upon treatment of iCell Neurons with increasing concentrations of paclitaxel (0.001-100 \( \mu M \)) for 72 h, we identified a reproducible 3–5 \( \mu M \) (12–14%) decrease in cell median neurite process length and a 13–19 \( \mu M \) (10–13%) decrease in cell total neurite outgrowth per order of magnitude increase in drug. Paclitaxel binds to beta-tubulin to exert its chemical effect. As a proof of concept, we have shown that decreased expression of the beta-tubulin isoform TUBB2A by siRNA transfection causes decreased median neurite process length (interaction \( P < 10^{-3} \)) and decreased total neurite outgrowth (interaction \( P < 10^{-3} \)) of iCell Neurons 48 h post-paclitaxel treatment. This work was supported by NIH grants U01GM61393 and F32CA165823.
718W Genetic variants associated with elevated triglyceride levels in patients with genotypes 2/3 chronic hepatitis C treated with cyclophilin inhibitor alisporivir. Y. Li1, M. Healey1, M. Waidvogel2, N. Hartmann2, G. Nabel1, L. Li1, F. Staudtler1, W. Zhang1, C.A. Brass1, N.V. Naoumova1, K.J. Johnson1, B. Li1. 1) Novartis Institutes for Biomedical Research, Cambridge, MA, USA; 2) Novartis Institutes for Biomedical Research, Basel, Switzerland; 3) Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA; 5) Novartis Pharmaceuticals Corporation, Basel, Switzerland.

Background: Host-targeting antiviral alisporivir (ALV) inhibits cyclophilin A that is essential for HCV replication. In the phase IIb VITAL-1 study involving treatment-naive genotype 2/3 patients with chronic hepatitis C, ~81% patients achieved SVR24 with IFN-free or IFN-containing ALV treatment with or without ribavirin. In a proportion (8.8%) of patients treated with ALV IFN-free regimen, maximum fasting triglyceride (TG) level of >400mg/dL was detected over the course of 24 weeks of treatment. The goal of this investigation was to evaluate the contribution of genetic variants to elevated TG levels observed in some patients in this study. Methods: DNA samples were obtained from 186 patients who consented to pharmacogenomic assessment and received ALV-containing treatment. Samples were genotyped using the Illumina OmniExome array and TaqMan allelic discrimination assays for APOE genotyping. Linear regression was performed to evaluate association between genotype and log-transformed maximum on-treatment TG level. The analysis was adjusted for race, log-transformed baseline TG, ALV exposure and presence of pegIFN treatment. Results: Targeted analysis of APOE genotype found that the APOE e2 carriers had higher levels of TG than non-carriers, in patients treated with ALV (p = 0.035). However, the frequency of e2 carrier was not associated with TG level (p = 0.12). Genome-wide association analysis (GWAS) with maximum fasting TG level while on ALV treatment identified one SNP in CNTNAP4 gene (in all races combined, p= 2.63x10^-9) and one SNP in THSD7B gene (in East Asians only; p=1.1x10^-6) that were associated with elevated level of TG, after adjusting for multiple testing, although functional relevance of these genes (CNTNAP4 and THSD7B) to level of triglyceride is unclear. Conclusions: Overall, these findings suggest a potential role of genetic variants in the level of TG in patients treated with ALV. Targeted analysis revealed a trend of association between elevated TG level and the presence of the APOE e2 allele in patients treated with ALV. Other genetic variations have also been identified from GWAS analysis of TG level in patients treated with or without ALV-containing therapy without clear functional interpretation. Testing of these candidate variants in independent studies is needed for validation.

719W Variants of the cysteinyl leukotriene 1 and 2 genes are additively associated with atopy in a founder population. M. Thompson1, J. Stankova2, V. Capra3. 1) Lab Medicine, Banting Inst, Univ Toronto, Toronto, ON, Canada; 2) University of Sherbrook, Sherbrook, QC, Canada; 3) University of Milan, Milan, Italy.

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

720W Identification of expression quantitative trait locus associated with drug biotransformation. H.-C. Yang1,2, C.-W. Chen1. 1) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; 2) School of Public Health, National Defense Medical Center, Taipei, Taiwan.

Scientific evidence has shown that drug response may vary with genetic background of populations. Pharmacogenomics association studies without accounting for population stratification and/or population admixture will cause false positive findings. Ancestry informative markers which exhibit different genetic distributions and reflect genetic ancestry in populations can be applied to adjust for population stratification and/or population admixture in genetics and pharmacogenomics association studies. This study aims to identify the ancestry informative markers which regulate gene expression and cause a down-stream consequence, differential drug responses in populations. We analyzed single nucleotide polymorphism and gene expression data of 210 independent samples in the HapMap II Project and identified a large number of cis- and trans-acting eQTL for each population by a partial least square method. Then we used our developed BIASLESS (Biomarkers Identification and Samples Subdivision) software (http://www.stat.sinica.edu.tw/hsinchou/genetics/prediction/BIASLESS.htm) [Yang et al., BMC Genomics, 2012, v13, 346] to identify ancestry informative eQTL that they are able to classify samples from different populations with a high testing accuracy in a cross-validation procedure. The identified eQTL were correlated with single nucleotide polymorphism on drug biotransformation genes. We found that the differential allelic distributions of cytochrome P450 enzymes, which are the essential enzymes involving in drug metabolism, in populations were found. The results provide an insight into the complex relationship of genetic ancestry, gene regulation and pharmacoresponse.

721W Genome-wide association study of combined paclitaxel and carboplatin treatment-induced severe neutropenia/leucopenia for patients in Biobank Japan. S. Loy1, S. Chung1, A. Takahashi1, T. Mushioda1, M. Kubo1, Y. Nakamura2,3. 1) Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Tokyo, Japan; 2) Department of Medicine, The University of Chicago, Chicago, IL; 3) Laboratory of Molecular Medicine, Human Genome Institute, Center of Medical Science, University of Tokyo, Tokyo, Japan; 4) Laboratory for Pharmacogenomics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 5) Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan.

Chemotherapeutic agents are notoriously known to have narrow therapeutical range that often resulted in life-threatening toxicity. Hence, it is clinically important to identify the patients who are at high risk for severe toxicity through pharmacogenomics approach. In this study, a genome-wide association study was performed with cancer patients who administered combined paclitaxel and carboplatin treatment to identify genetic variants that are associated with the risk of severe neutropenia/leucopenia in the Japanese population. A total of 477 patients’ DNA were recruited from the Biobank Japan with 161 patients who did not develop any drug reactions, 161 who develop mild (grade1/2) and 150 who develop severe (grade3/4) neutropenia/leucopenia after administering the combined treatment. All of these samples were genotyped using Illumina OmniExpress BeadChip™. After sample and quality control, two genetic loci, rs12310399 (P=2.46x10^−7, OR=1.85, 95% CI =1.32-2.58) at FGD6 gene and rs10785877 (P=7.38x10^-8, OR=2.58, 95% CI =1.77-3.77) near RXRA gene, were identified to be suggestive associated with paclitaxel and carboplatin treatment induced severe neutropenia/leucopenia when genome-association study was performed with patients who do not develop any adverse drug reaction versus those who develop severe neutropenia/leucopenia. Weighted genetic risk score analysis using six SNPs with P<1.0x10^-5 identified that individuals who carry the highest range of genetic risk score possess 188 times (95% CI =36.1-979) higher risk to develop severe neutropenia/leucopenia compared to patients who belong to the lowest range of genetic risk score. Interestingly, individuals who developed grade 1/2 (mild neutropenia/leucopenia) were found to show intermediate risk scores between patients with severe neutropenia/leucopenia and those without any adverse reactions. Although we failed to identify genetic variants that surpassed the genome-wide significance level (P≤5.0x10^-8) through GWASs probably due to insufficient statistical power and complex clinical features, we were able to shortlist some of the suggestive associations. This current study is at the relatively preliminary state, but could highlight the complexity and problematic issues in retrospective pharmacogenomics studies. However, we hope that verification of these genetic variants through local and international collaborations could improve the clinical outcome of cancer patients.
722W
Expression of MxA, OAS1, PKR and TP53 - interferon stimulated genes during treatment of hepatitis C patients. B. Świątek1, I. Bereszynska2, A. Kowala-Piaskowska1, J. Rembowska1, J. Nowak2. 1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) University of Medical Sciences, Poznań, Poland; 3) Department of Medical Diagnostics, Poznan, Poland.
Interferon stimulated genes (ISGs) such as MxA, OAS1, PKR play a key role in antiviral responses against HCV infection. Moreover, it is suggested that ISGs pre-activation is associated with anti-HCV treatment failure. Also, it was observed that interferon stimulates transcription of TP53 gene. The aim of this prospective study was to examine the association between MxA, OAS1, PKR (EIF2AK2) and TP53 expression and response to pegylated interferon and ribavirin treatment in 35 chronic hepatitis C (CHC) patients. Viral load was determined using one-step quantitative RT-PCR. Analyses were performed before as well as at 4 and 12 week of treatment. Rapid Virological Response (RVR) and complete Early Virological Response (cEVR) was achieved by 33 (77.8%) and 10 (23.9%) patients, respectively. 12 (34.3%) did not respond to pegIFN and ribavirin combination treatment during 12 weeks (Non-Response, NNR, less than 2 log10 decrease in viral titer after 12 weeks of therapy). The mean baseline viral load was comparable in RVR, cEVR and NNR group (6.7, 7.3 and 3.5±104 IU/ml, respectively). Expression of classical ISGs (MxA, OAS1, PKR), but not TP53 increased during CHC treatment. The expression was low in RVRs and higher in cEVRs and NNRs before therapy and increased noticeable in cEVRs and NNRs and poorly in RVRs at week 4. The expression of studied ISGs was stable or poorly decreased in RVRs, was stable or poorly increased in cEVRs and strongly increased in NNRs. The high level of ISGs tested might predict the outcome of combination CHC treatment. It looks like, pre-activation of the endogenous interferon system is associated with RVR and thereby with high likelihood of achieving SVR. It is also possible that treatment failure during first 12 weeks of therapy (PNR) may be related to noticeable decrease in ISGs expression between weeks 4 and 12. Lastly, it can be suggested that there is no association between TP53 expression during interferon treatment of CHC patients. This research is supported by National Science Centre grant no 2011/01/B/N26/04258.

723W
Analysis of the CYP3A4, CYP3A5, CYP2C9, CYP2C19 and CYP2D6 polymorphism in Nayarit population: Involvement in drug metabolism. LE. Wong-Ley, AB. Martínez Rizo, JB. Velázquez Fernández, JM. Wysokić1, M. Mozor-Liszewska1, J. Rembowski2, J. Nowak2. 1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) University of Medical Sciences, Poznań, Poland; 3) Department of Medical Diagnostics, Poznan, Poland.
Genetic variants in the cytochrome P450 (CYP) genes, contribute to pharmacological, therapeutic and toxicological outcomes. These genetic variants are associated with changes in the hepatic transporter or with their own metabolism. They can have significant implications for drug response and toxicity, as well as for diagnosis and treatment options. Numerous studies have shown that some CYP polymorphisms may influence the pharmacokinetics and pharmacodynamics of drugs, particularly those that are metabolized by the liver. The aim of this study was to investigate the prevalence of the CYP3A4, CYP3A5, CYP2C9, CYP2C19 and CYP2D6 polymorphisms in the Nayarit population, a West Mexican state with polymorphisms in genes encoding them, so it is likely that these polymorphisms are involved in the response to these drugs. In Mexico there are few studies of the allelic frequencies of these genes and their correlation with the treatment of various diseases. The objective the present study is to know the association between the CYP3A4*1B, CYP3A5*3, CYP2C9*1, CYP2C9*2, CYP2C19*1, CYP2C19*2 polymorphisms and the response to therapy of various drugs in Nayarit, a West Mexican population. They were included in the study 200 healthy patients. All patients signed an informed consent and a questionnaire was answered. DNA was obtained from lymphocytes and followed the identification by PCR and enzymatic digestion of CYP3A4*1B, CYP3A5*3, CYP3A5*6, CYP2C9*2, CYP2C9*3 and CYP2C19*2 variants. All genes studied are in Hardy-Weinberg equilibrium. The allelic frequencies of the studied genes were: CYP3A4*1B, CYP3A5*3, CYP3A5*6, CYP2C9*2, CYP2C9*3 and CYP2C19*2 genes in our state have relatively high frequencies compared to other populations, such as Caucasians, Asian or Africans, indicates a strong relationship between therapeutic response shown in our population and the genetic polymorphisms found, which can be used to develop personalized treatments for better therapeutic management of diseases. We believe that this will be possible through better understanding of the physiology of both, the disease and the patient, thus allowing for personalized treatments.

724W
Can IL28B polymorphism identify patients who achieve early virological response during treatment of hepatitis C? J. Nowak1, B. Świątek1, I. Bereszynska2, A. Kowala-Piaskowska1, M. Mozor-Liszewska1, J. Rembowski2, J. Januszkiewicz-Lewandowska1,2,3, 1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) University of Medical Sciences, Poznań, Poland; 3) Department of Medical Diagnostics, Poznan, Poland.
Large number of factors affecting response to hepatitis C therapy makes it difficult to optimize. Identification of molecular markers playing role in predicting anti-HCV treatment outcome before its beginning would facilitate individualization of HCV therapy. Previously, the IL28B rs12979860 C/T polymorphism has been identified as a strong predictor of Sustained Virological Response (SVR) in chronic hepatitis C (CHC) patients treated with pegylated interferon and ribavirin (pegIFN-RBV). Moreover, having a better understanding of the biology of these polymorphisms may help identify patients with high chances of treatment success. The aim of this prospective study was to examine the association between SNP rs12979860 (C/T) in the IL28B gene and on-treatment virological response (week 4 and 12) in CHC patients. The study consisted of 35 patients who were treated with pegylated interferon (PEG-2A) and ribavirin. To determine treatment effects, serum HCV-RNA was measured on the first day of therapy and then at 4 and 12 week of therapy by one-step quantitative RT-PCR. DNA, isolated from peripheral blood lymphocytes, was used for IL28B rs12979860 (C/T) genotyping by High Resolution Melting method. 13 patients (37.1%) became HCV RNA negative at week 4 (RVR-Rapid Virological Response) and 10 (28.6%) at week 12 (cEVR-complete Early Virological Response). 12 patients (34.3%) did not achieve virological response until 12 week of therapy (PNR). Among patients with CC genotype, RVR was observed in 30.5%, 30.5% and 39% of patients, respectively. 25% of patients with the genotype CT were observed in RVR, cEVR and PNR group. Moreover, increase in ISGs expression at week 4 of CHC therapy might differ between those who achieve RVR, cEVR and PNR. The rs12979860 CC, CT and TT genotyping was stable or poorly increased in RVR, cEVR and PNR group, respectively. Among patients with CC genotype, 75% achieved RVR and 25% achieved cEVR. Among CT genotype RVR, cEVR and PNR were observed in 30.5%, 30.5% and 39% of patients, respectively. 25% of patients with the genotype TT achieved RVR and 75% achieved cEVR. Favorable CC was not observed in RVR and unfavorable TT genotype was not observed in RVR group. The initial results confirm that IL28B rs12979860 C/T polymorphism may identify those CHC patients, treated with pegIFN and ribavirin, who achieve virological response during early phase of treatment and thereby who are likely to achieve SVR. This research was supported by National Science Centre grant no 2011/01/B/N26/04258.

725W
Effect of genetically tailored statin therapy on health behaviors and outcomes: A pilot study in the primary care setting. J.H. Li1, S.V. Joy2, S.B. Haga1, L.A. Orlando1, W.E. Kraus1, G.S. Ginsburg1, D. Voora3. 1) Duke University Medical Center, Durham, NC; 2) University of Colorado Denver Anschutz Medical Campus, Aurora, CO; 3) University of Paris Descartes, Paris, France.
Despite the cardiovascular benefits of statins, long-term adherence is often limited by real or perceived side effects. The *5 variant in the hepatic transportor SLC01B1 is a risk factor for myopathy, a common side effect of statins. Our study aimed to examine the simultaneous effect of statin and atorvastatin (S/A) and least with pravastatin and rosuvastatin (P/R). We hypothesized that providing SLC01B1*5 genotype guided statin therapy (GGST) would be associated with improved patient adherence, provider behavior, and laboring-on outcomes in primary care patients with chronic hepatitis C. We recruited 35 patients who were initially nonadherent to statins in a pilot study. Methods: Patients (n=58) and their providers received *5 genotype results and GGST recommendations pushed to them via the electronic medical record (EMR). Noncarriers were given increased risk of side effects and were reassured to restart any statin, while carriers had a higher risk of side effects from S/A and were recommended therapy with P/R. The primary outcome was the change in patients' perceived need for statins and concerns about risks from baseline to 12 months, measured by a validated survey. Concurrent controls (n=59) receiving standard of care were gathered from the same clinic for comparison of secondary outcomes: 1) proportion with new statin prescriptions written by their providers, 2) change in low-density lipoprotein cholesterol (LDLC), 3) patient-reported statin usage over 1 year. Results: The largest changes were in the need for statin to prevent sickness (p=0.01, odds ratio = 3.3±0.9, p<0.001) and the ‘concern for statin to disrupt life’ (p<0.001). Overall, GGST patients expressed a trend toward higher necessity (p=0.02) and lower concern (p=0.004) than patients on statins. GGST patients had a higher proportion of new statin prescriptions (55% vs. 20%, p=0.001) and greater change in LDLC (-1.7±0.48 vs. 6.3±3.78, p=0.003) and higher patient-reported statin use (45% vs. 15%, p=0.001) during follow-up. Conclusion: Delivery of SLC01B1*5 genotype guided recommendations to tailor statin therapy is feasible via the EMR in the primary care setting. This novel intervention improved patients’ perceptions of statins and was associated with physician and patient behaviors that promoted higher statin adherence and lower LDLc. The impact of *5 genotyping should be further explored in a larger randomized control trial.
Pharmacogenetics

Incidental pharmacogenetic variants identified by massively parallel sequencing in the ClinSeq® study. D. Ng1, J.J. Johnston1, K.L. Lewis1, S.G. Gospe1, W. Nation2, S. Crosslin3, D.N. Crosslin3, D.A. Nickerson3, J.C. Mulhkin5,5, L.G. Biesecker1,4, NIH Intramural Sequencing Center, 1) Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Manchester Centre for Genomic Medicine, St. Mary’s Hospital, Manchester, UK; 3) Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, UK; 4) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Genome Technology Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Massively parallel sequencing can identify incidental, medically relevant genetic variants in asymptomatic individuals. This represents a major paradigm shift in medicine, from the treatment of manifest disease to predictive medicine aimed at monitoring asymptomatic individuals with disease susceptibility. One of the aims of ClinSeq® is to study the analysis and return of incidental, clinically relevant genetic variants (secondary findings). As part of our ongoing endeavor to study secondary variants, we selected 20 genes associated with drug metabolism or transport for analysis. Whole exome sequences (WES) from 951 ClinSeq® participants were annotated for variants in 20 drug metabolism genes selected from the Affymetrix DMET® plus array using the Human Gene Mutation Database (HGMD®) cDNA/protein reference. Of the 327 variants targeted by the Affymetrix DMET® plus chip, 70.6% (231/327) had sequence coverage in ≥80% (761–951) and 20.5% (67/327) were covered in <80% (1–759) of participants. WES identified 360 variant positions in the 20 genes. Ninety-two of the pharmacogenetic variants that were identified by the Affymetrix chip were identified by WES in one or more ClinSeq® participants. Twenty-four hundred sixty-eight variant positions were novel. Twenty-four of these variants were predicted to cause a loss of function (frameshift, nonsense, splice). The remainder of the 244 coding changes were missense variants without a reported phenotype. The 82 variants identified by WES, and targeted by the Affymetrix chip were categorized in HGMD® as follows: functional polymorphism (n=66), disease-associated polymorphism (n=8), disease-associated polymorphism with additional supporting functional evidence (n=11) and disease-causing (n=7). Variants with potential medical relevance include VKORC1 p.Asp36Tyr reported to predispose to warfarin resistance (n=5 individuals), SLC01B1 p.Val174Ala associated with higher basal cholesterol (n=20 individuals) and CYP2C8 p.Leu68Met associated with an increased risk of coronary artery disease among smokers (n=67 individuals). Three hundred and sixty variants were identified in 20 genes associated with drug metabolism. WES identified insertion/deletion variants, but not promoter or deep intronic SNPs. Ninety-two variants have published data suggestive of a role in drug metabolism. Further research is needed to determine which of these variants have clinical utility thereby merit the return of genetic results to participants.


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

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

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


Molecular inversion probes (MIPs) are flexible, customizable, scalable and affordable method to perform massively parallel target capture and sequencing with a minimal amount of genomic DNA input. With the intention of streamlining and scaling a MIP capture platform to process thousands of samples with any given panel of gene candidates, we characterized the capture efficiency of a 10-gene panel of pharmacologically significant genes that contain a spectrum of challenging capture targets; highly paralogous, repetitive and GC-rich (i.e., CYP2D6) to unique sequence (i.e., COMT). All MIPs in our 10 gene panel were pooled at equimolar ratios and captured at an initial genomic DNA target to MIP ratio of 1:200. Due to the complexity of the gene panel, we observed non-uniform MIP capture across our targets requiring rebalance of the MIP pool to increase the likelihood of capture for inefficient MIPs. To eliminate the need for time-consuming successive pool rebalancing of individual samples, we assessed capture repeatability in two ways: intra-sample (using 12 replicates of a single sample) and inter-sample (across four HapMap trios that represent diverse ancestry and CYP2D6 genotype). Pairwise comparisons between capture reactions from both repeatability assessments show high correlation. Utilizing the normalized MIP capture events, we developed a method to rebalance the 10 gene panel MIP pool to compensate for poorly performing probes. We observed that MIP performance correlates with the nature of the gene complexity and the sample analysis. A tool named FAMOUS was used to include feedback from targeted MIP pool to serve as a capture efficiency control, for sample identity confirmation and to provide quality control monitoring for sample contamination. In summary, we have optimized the MIP capture platform by eliminating successive pool rebalancing, and developing reagents, and incorporating robotic for making and rebalancing probe pools to increase throughput.
730W
Assessment of the predictive effect of genetic variation in key genes associated with drug therapy of cardiovascular diseases in the Azores Islands (Portugal), L. Mota-Vieira1,2,3, M.S. Melo1,4, L. Balcano1,2, R. Cabral1, C.C. Branco1,2,3, M. Pereirinha1, 1) Molecular Genetics and Pathology Unit, Hospital of Divino Espírito Santo of Ponta Delgada, EPE, Azores, Portugal; 2) Instituto Gulbenkian de Ciência, Lisbon, Portugal; 3) BioFig, Lisbon, Portugal; 4) The University of the Azores, Ponta Delgada, Portugal.

New understandings of individual genetic variants are increasingly reshaping, in various aspects, the drug treatment of cardiovascular diseases (CVD). Three highly prescribed drugs - clopidogrel (platelet aggregation inhibitor), warfarin (anticoagulant) and simvastatin (cholesterol-lowering drug) - are examples with strong genetic evidence. Here, we investigate the predictive effect of relevant pharmacogenes - CYP2C19 (clopidogrel bleeding risk), CYP2C9 and VKORC1 (warfarin bleeding risk), and SLC01B1 (simvastatin-induced myopathy risk, from myalgia to rhabdomyolysis) - associated with drug response in Azoreans. This population could benefit from these pharmacogenetic data, since CVD are the first cause of mortality and morbidity in Azores Islands. According to SNP information available on PharmGKB, Pubmed, dbSNP, and SNPedia databases, DNA of 170 blood donors was genotyped by TagMan® genotyping assays (Applied Biosystems). Our results demonstrate that in Azoreans the influence of the CYP2C19 loss-of-function alleles (‘2’ and ‘3’) on clopidogrel response is only due to the ‘2’ allele. Therefore, clinicians should prescribe a non-CYP19 dependent thienopyridine, such as prasugrel or ticagrelor, in homozygous (1.8%) and heterozygous (27.6%) genotypes, respectively. Regarding warfarin, the joint analysis of CYP2C9 (‘2’ and ‘3’) and VKORC1-3 revealed that 82.4% of Azoreans will need intermediate or low drug doses, guided by a clinical algorithm, if treatment is started. In what concerns the SLC01B1 rs5217>C SNP, known to cause high plasma levels of simvastatin, three genotypes were observed: TT (individuals with two functional alleles, 70.6%), TC (one functional allele plus one reduced-function allele, 27.6%) and CC (two reduced-function alleles, 1.8%). Thus, clinicians should be alert to simvastatin-induced myopathy in patients with intermediate and high risk (TC and CC, respectively) and (whenever needed), prescribed dose adjustments are required. Further studies to validate uncharacterized ADME variation (including SNVs, indels, and CNVs) are being performed. Ultimately, these results will help enhance understanding of ADME gene variation across ethnic groups which has not been captured in existing genotyping panels and which warrants further functional characterization.

731W

Identifying factors which influence the pharmacokinetics of a drug (the body’s exposure to the drug) and selecting optimal doses is a crucial part of drug development. Variation in genes that regulate drug absorption, distribution, metabolism and excretion (ADME) can contribute to variability in drug pharmacokinetics; in addition, the frequency of variants in ADME genes may differ between ethnic groups. Despite this, there have been few systematic studies to identify and characterize the extent of variation in ADME genes in different ethnic groups. To address this, we called single nucleotide variants (SNVs), short insertion-deletions (indels) and gene copy number variants (CNVs) in exome sequence data from healthy Northeast Asian (NEA) subjects (Korean N=126; Japanese N=125) and subjects from 19 different ethnic groups in the 1000 Genomes Project (1000G; N=1181). Average depth across the core (N=38) and extended (N=266) ADME genes (as defined by PharmaADME.org) is ~60X and ~130X for the NEA and 1000G subjects, respectively. Approximately 75% of ADME coding bases were well-captured (sequenced to an average depth >20X) in the combined sample (NEA+1000G). NEA subjects carried on average approximately 180 nonsynonymous (NS; 15% predicted to be deleterious) and 9 loss-of-function SNVs across all ADME coding regions. On average, each NEA subject carried ~9 NS ADME coding SNVs currently uncharacterized in public databases of functional ADME variation, of which 90% are predicted to be deleterious. NEA subjects carried on average ~1 frameshift indel in ADME coding regions and exhibit evidence of potentially novel disruptive gene deletions or duplications in UGT and CYP gene families. Restricting focus to well-captured ADME regions across the combined sample, NEA subjects carried slightly fewer NS variants (~135 SNVs per person) than 1000G subjects (~141 SNVs per person, ~35% European (EUR) and ~174 African (AFR)). Quantities of uncharacterized NS ADME variation predicted to be deleterious were similar in the NEA, ASN and EUR subjects (2.0, 2.2 and 2.0 per person, respectively) and higher for the AFR subjects (4.5 per person). Follow-up experiments to validate uncharacterized ADME variation (including SNVs, indels, and CNVs) are being performed. Ultimately, these results will help enhance understanding of ADME gene variation across ethnic groups which has not been captured in existing genotyping panels and which warrants further functional characterization.
732W
The CYP2C19*17 Variant is not Independently Associated with Clopidogrel Response. J. Lewis1,2, S. Stephens1,2, R. Horeinstein1,2, J. O'Connell3, K. Ryan1, C. Peer4, W. Figg5, S. Spencer6, M. Pacanowski6, B. Mitchell3, A. Shuldiner1,2,6 Endocrinology, Diabetes & Nutrition, University of Maryland, Baltimore, Baltimore, MD; 2) Program in Personalized and Genomic Medicine, University of Maryland, Baltimore, Baltimore, MD; 3) Clinical Pharmacology Program, National Cancer Institute, Bethesda, MD; 4) Applied and Developmental Research, SAIC-Frederick Inc., National Cancer Institute, Frederick, MD; 5) U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Office of Clinical Pharmacology, Silver Spring, MD; 6) Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, MD.

Background: Cytochrome P450 2C19 (CYP2C19) is the principle enzyme responsible for converting clopidogrel into its active metabolite and genetic variants have been identified, most notably CYP2C19*2 and CYP2C19*17, that are believed to alter its activity/expression. We evaluated whether the consequences of the CYP2C19*2 and CYP2C19*17 variants on clopidogrel response were independent of each other or genetically linked through linkage disequilibrium (LD). Methods: We genotyped the CYP2C19*2 and CYP2C19*17 variants in 621 members of the Pharmacogenomics of Anti-Platelet Intervention (PAPI) Study and evaluated the effects of these polymorphisms singly then jointly, taking into account LD on clopidogrel prodrug level, clopidogrel active metabolite level, and ADP-stimulated platelet aggregation pre- and post-clopidogrel exposure. Results: The CYP2C19*2 and CYP2C19*17 variants were in LD (iD=1.0; r2=0.07). In association analyses that did and did not account for the effects of CYP2C19*2, CYP2C19*17 was strongly associated with levels of clopidogrel active metabolite (beta=-5.24, P=3.0x10^-10 and beta=5.36, P=3.3x10^-14, respectively) and post-treatment ADP-stimulated platelet aggregation (beta=7.55, P=2.9x10^-16 and beta=7.51, P=7.0x10^-15, respectively). In contrast, CYP2C19*2 was associated with clopidogrel active metabolite levels and ADP-stimulated platelet aggregation before (beta=5.77, P=1.7x10^-11 and beta=9.98, P=1.1x10^-14, respectively) but not after (beta=0.40, P=0.59 and beta=0.13, P=0.69, respectively) adjustment for the CYP2C19*2 variant. Stratified analyses of CYP2C19*2/CYP2C19*17 genotype combinations revealed that CYP2C19*2 was a primary determinant in altering clopidogrel response. Conclusions: Our results suggest that CYP2C19*17 has a small (if any) effect on clopidogrel-related traits and that the observed effect of this variant is due to LD with the CYP2C19*2 loss-of-function variant.

733W

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

734W
Exome sequencing of asthmatics with extreme corticosteroid response. Q.L. Duan, M.H. Cho, R. Kelly, W. Qu, E.R. Mardis, E.R. Bleecker4, D.A. Meyers3, G.A. Hawkins5, S.P. Peters2, J.J. Lima2, K.G. Tantisira1, S.T. Weiss1,2. 1) Channing Div Network Med, Brigham & Women’s Hosp, Boston, MA; 2) Pulmonary Division, Brigham & Women’s Hospital and Harvard Medical School, Boston, MA; 3) Partners Center for Personalized Genetic Medicine, Partners Health Care, Boston, MA; 4) The Genome Institute at Washington University School of Medicine, St. Louis, MO; 5) Center for Genomics and Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, NC; 6) Center for Pharmacogenomics and Translational Research, Nemours Children’s Clinic, Jacksonville, FL.

Corticosteroids (CS) are potent, anti-inflammatory drugs used for the treatment of numerous diseases including asthma, which affects over 300 million individuals worldwide and 20 million Americans. However, great inter-individual variability in response to this class of drugs has been described, whereby up to 25% of patients are non-responsive. Previous pharmacogenetic studies of CS response focused primarily on the contribution of common polymorphisms, whereas increasing evidence suggests that rare variants are important genetic determinants of complex traits and confer larger effect sizes. We hypothesize that novel rare variants, identified through next-generation sequencing of whole exomes, contribute to the heterogeneity of CS response in asthma patients. To test this, we selected 196 individuals including equal numbers of poor and good responders to inhaled CS, measured as ≥50% and ≥13% change in FEV1, forced expiratory volume in one second) following 4 to 8 weeks of therapy. Exome sequencing was performed using the NimbleGen SeqCap EZ v2.0 array on the Illumina HiSeq 2000. Sequences were aligned using BWA, and variants called using Samtools and VarScan. After removing duplicates and mono- and biallelic variants (SVs), of which 347,802 passed quality control in 188 individuals (91 poor and 97 good responders). Of these, 75,206 coded for non-synonymous, splice, or stop variants, the majority of which were rare (60,725 or 81% with MAF < 0.05). Our primary analysis in SKAT-O was a per-gene burden test of these rare variants, weighted by MAF, and adjusted for gender, age, and height. Our top genes (P-values between 2.8 to 9.2x10^-13) were FBXO6, CD163, D2RB2, ARHGEF16, PCK1, FOXK1, CASC5. Although none of these were exome-wide significant, our secondary analyses identified several of the top loci are differentially expressed in response to dexamethasone (a corticosteroid) in the lymphoblastoid cell lines of asthmatics. Furthermore, CD163 and PCK1, are excellent biological candidates that are known to be regulated by glucocorticoids. In conclusion, exome sequencing of extreme CS response in asthmatics may identify novel loci that contain rare variants that could predict a CS response phenotype. A secondary analysis using illumina Exome Chip data in nearly 400 asthma trios is under way.
Two-Hit Models of Disease Risk in Drug-Induced Liver Injury. J.I. Goldstein1,2, E.T. Lim1,2,3, H. Huang1,2, S. Raychaudhuri1,4,5, C. Stevens4, P.I.W. de Bakker1,2,7,6, M.I. Lucena5, A.K. Daly10, M.R. Nelson11, A. Hokken12, B.M. Neale1,2, M.J. Daly1,2, International Severe Adverse Events Consortium. 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Division of Genetics, Brigham and Women’s Hospital, Boston, MA; 5) Division of Rheumatology, Immunology, and Allergy, Brigham and Women’s Hospital, Boston, MA; 6) Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA; 7) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; 8) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 9) Hospital Universitario Virgen de la Victoria, Malaga University, IBIMA and Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Spain; 10) Institute of Cellular Medicine, Newcastle University Medical School, Newcastle Upon Tyne, UK; 11) Quantitative Sciences, GlaxoSmithKline, Research Triangle Park, NC; 12) International Severe Adverse Events Consortium, Ltd., Chicago, IL.

Genome-wide association studies for drug-induced liver injury (DILI) have found strong associations with large effect sizes to classical human leukocyte antigen (HLA) alleles. However, the frequencies of the risk alleles in the treated population are three orders of magnitude higher than the prevalence of DILI (approximately 8/100,000), suggesting a role for other genes in DILI. It has been proposed that part of the missing heritability in complex diseases can arise from non-additive genetics, such as epistasis. Therefore, we hypothesize that there might be a two-hit model involved in conferring risk for DILI. To test this hypothesis, we genotyped samples with two DILI phenotypes (fluocloxacillin DILI: n=69; co-amoxiclav DILI: n=124) and ancestry-matched controls (n=3159) on the Illumina HumanExome array. When we imputed the classical HLA alleles and performed single-variant association tests, we obtained similar results to what had previously been reported (fluocloxacillin: HLA-B*57:01 (p=3e-35, OR=50); co-amoxiclav: HLA-DQB1*06:02/HLA-DRB1*15:01 (p=3e-18, OR=3.4)). However, single-variant tests for rare variants did not yield any significant associations, suggesting we are underpowered to detect rare risk variants with moderate effect sizes. To test the two-hit model of disease, we calculated binomial probabilities of observing an excess of cases with rare alleles at two unlinked genes and then compared them to empirical p-values using permutation testing to control for population stratification. We also implemented a conditional scanning approach by incorporating what is known about the biology of DILI from the literature. For example, Monshi et al. showed fluocloxacillin covalently binds to lysine residues in albumin and three of these residues occur in amino acid sequences that are predicted to bind to HLA-B*57:01 (Hepatology, 2013). Therefore, we looked for sets of variants that mimicked this scenario in other genes expressed in the liver and found three variants that look promising for fluocloxacillin DILI in CYP2A5, PMO1, and CD68. In this work, we demonstrate novel approaches for testing a two-hit model in a complex disease and present results that demonstrate the utility of testing such a model in pharmacogenetic studies.

Pharmacogenomics of glucuronidation in American Indians: The Strong Heart Family Study. PE. Melton1, N. Franceschini2, K. Haack2, C. Bizon2, ET. Lee5, JG. Umans3, LG. Best2, SA. Cole3, LA. Alam32, 1) Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley, Western Australia, Australia; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 4) Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC; 5) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 6) Medstar Health Research Institute, Hyattsville, MD; 7) Missouri Breaks Research Industries, Inc, Timber Lake, SD.

Glucuronidation is a phase II (conjugation) process facilitating both biliary and renal clearance of both drugs and endogenous substances. Bilirubin serves as an endogenous glucuronidation biomarker, as its levels are affected by the UDP glucuronosyl transferase (UGT) and organic anion transporter (OAT) products encoded by the genes UGT1A1 and SLC01B1, respectively. Polymorphisms in these genes are also known to impact clearance of several cardiovascular (CVD), and antineoplastic drugs. Previous research on genes influencing bilirubin in American Indians from the Strong Heart Family Study (SFHS) detected a QTL (LOD=6.61) on chromosome 2q near the UGT1A1 gene, but linkage conditional on an associated repeat promoter polymorphism, rs5839491, did not completely explain the linkage signal. Genetic testing of this particular variant is used routinely to guide dosing of irinotecan, an antineoplastic agent. To elucidate additional functional variants influencing glucuronidation and bilirubin metabolism, we conducted whole exome sequencing (Illumina TruSeq Exome Enrichment Kit) in SHFS participants from Arizona (n=47) and Oklahoma (n=47). Genetic diversity was maximized by selecting founders from distant relatives (Arizona) or by principal component analysis (OK). Average sequence coverage was 36x and 263,378 variants passed quality control. Using a priori evidence we identified 398 variants in 31 genes from the UGT family and 7 genes involved in bilirubin metabolism. Of these variants, 83 were non-synonymous and 8 were novel. These included the UGT1A1 exon variant, rs4148323, which was detected in 4% of sequenced SFHS participants and was significantly associated with total bilirubin (p=0.0004). This SNP is known to influence bilirubin metabolism and is associated with increased irinotecan toxicity in East Asian populations but is absent in Europeans. We also detected the exonic SLC01B1 SNP rs4149056 in 4% of these SHFS participants, which is known to impact statin metabolism, effect, and toxicity. This variant was not significantly associated with total bilirubin (p=0.86) in these SHFS participants. These results demonstrate the efficacy of whole exome sequencing to identify novel pharmacogenetic markers in American Indians and may have implications for recommended testing to decrease adverse drug events in this population.
Primary action of clozapine exposure on activation of SREBP-controlled lipogenic gene expression may explain benefit and detriment. J.S.A. de With 1,2, T. Wang 3, E. Strengman 4, S. de Jong 5, W.G. Staal 6, R.A. Ophoff 7, 1) Donders Institute for Brain, Cognition and Behavior, Department of Cognitive Neuroscience, Radboud University Medical Center Nijmegen, Utrecht, Netherlands; 2) UCLA Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, California, USA; 3) Karakter, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands.

Clozapine is an atypical antipsychotic drug with superior efficacy in the treatment of treatment resistant schizophrenia. However, its use is complicated by metabolic adverse effects, weight gain and clozapine-induced agranulocytosis (CIA), a potential lethal adverse effect occurring in approximately 1% of patients. This study aimed to investigate the mechanism behind the (adverse) effects of clozapine by using high throughput expression profiling and the genetic background of response variation in clozapine exposed lymphoblast cell lines. We performed two experiments to study the molecular basis of clozapine exposure at the cellular level. First, we established gene expression profiles of four lymphoblast cell lines that were exposed to a range of concentrations of clozapine (2µM to 100µM) for 24h. Secondly, we exposed 90 lymphoblast cell lines (HapMap CEU trios) to a range of concentrations of clozapine (2µM to 140µM) for 48h to create survival curves. The survival for each cell line is used as a quantitative trait measure for genetic mapping. Available gene expression data of these cell lines will be used to gain insight in molecular mechanisms behind differences in cell survival in response to clozapine exposure. We observed significant gene expression changes after exposure to clozapine in lymphoblast cell lines. Using a supervised approach, we found several genes that are significantly associated with clozapine exposure. Gene enrichment analyses showed very strong activation of Sterol Regulatory Element Binding Protein (SREBP) together with SREBP-target genes involved in sterol biosynthesis. Cell survival for lymphoblast cell lines showed individual differences suggestive of genetic factors playing a role in clozapine response, which may be important for identification of agranulocytosis susceptibility loci; genetic analysis is ongoing. The observed gene expression upregulation of genes from the SREBP pathway is a recurrent theme in different cell types and is also known to be important for identification of agranulocytosis susceptibility loci; genetic analysis is ongoing.

Objective: We will focus on the Genetic Testing Registry (GTR) and ClinVar. The GTR is a centralized, international registry of genetic test information covering clinical diagnosis and genetic testing for disease susceptibility, Nashville, TN. The GTR is a primary resource for locating detailed, practical information about pharmacogenetic tests and the indications for which tests are performed (www.ncbi.nlm.nih.gov/gtr). As of June 2013, laboratories have voluntarily provided information for 14 pharmacological responses (e.g., abacavir, clopidogrel, warfarin). Registered tests describe target patient population, indications, gene variant targets, and ordering details. NCBI staff work with authoritative groups including the Pharmacogenomics Knowledge Base (PharmGKB), and the Clinical Pharmacogenetics Implementation Consortium (CPIC). Customized links are provided to practical information in PharmGKB such as dosing guidelines and clinical annotations of gene variants. When available, succinct, curated synopses of drug responses by CPIC are presented. Practice guidelines from CPIC, the Pharmacogenetics Working Group of the Royal Dutch Association for the Advancement of Pharmacy (KNMP), and other authorities facilitate the translation of genetic test results into clinical actions for specific drugs. GTR also links to the FDA labels from NLM’s DailyMed. The lack of structured, open-access, peer-reviewed, comprehensive, gene-drug information is a deterrent to clinical implementation. To fill this gap, NCBI collaborates with CPIC and other experts to create Medical Genetics Summaries, available from NCBI’s Bookshelf (http://www.ncbi.nlm.nih.gov/books/NBK213999). These summaries are established with extensive literature cross-referencing with links to PubMed and OMIM. GTR also provides information about the genes involved in the drug response and variants that affect drug metabolism. Clinicians and researchers can access ClinVar (www.ncbi.nlm.nih.gov/clinvar) for lists of clinically relevant variants observed in humans, their phenotypic relationships, current interpretations, and supporting evidence.

Pharmacogenetic (PGx) studies require genetic testing of individuals for multiple variants in drug metabolism enzyme and transporter genes. For phenotyping and interpretation purposes, genotyping results must be translated to ‘star allele’ nomenclature. Star alleles are haplotype patterns that have been defined at the gene level and, in many cases, associated with protein and drug activity. Genetic variants within a haplotype can include SNPs, InDelS, and copy number variants. Knowing the combination of variants within a given haplotype, and the diplaid content in an individual, is of key importance for studying drug metabolism, drug response and adverse drug reactions. To facilitate the translation of results for individuals genotyped in studies using TaqMan™ Drug Metabolism Genotyping Assays and TaqMan™ Copy Number Assays, we developed a web-based flexible software tool called AlleleTyper™. This software uses genetic pattern information in user-defined translation tables to map sample genotyping data to star allele nomenclature. Typically, haplotype information in public resources such as the Cytochrome P450 Allele Nomenclature or PharmGKB databases is used to create a translator for gene alleles of interest. A translator can include one or more genes. A monallelic translation table is first prepared that contains the star allele/haplotype pattern information for each TaqMan assay in the study. This is imported into AlleleTyper, which automatically converts it to a biallelic translator containing diploid genetic pattern information. The biallelic translator is reviewed and edited, if needed. Subsequently, a study is created in AlleleTyper, the biallelic translator is imported, and then genotype results that have been mapped to TaqMan Genotyping Software (version 5.0) are imported. Other genotyping Software are also imported. AlleleTyper matches the genetic information in these files to the patterns in the biallelic translator and reports the star allele genotypes determined for each individual. The software includes error reporting for missing data, unmatched patterns, etc. A review of the software workflow and features will be presented, along with data analysis examples. AlleleTyper Software greatly facilitates PGx study data analysis, particularly for high throughput studies. This software is also flexible enough to be used for other genotyping applications requiring translation of data from multiple TaqMan assays, including triallelic SNP data analysis and blood genotyping.

742W Genetic ancestry modifies effects of naltrexone on smoking cessation in African Americans: an analysis of a randomized controlled trial. A.P. Bress1, C. Wing2, A. King3, R. Kittles4. 1) Pharmacy Practice, University of Illinois at Chicago, IL; 2) Health Policy and Administration, University of Illinois at Chicago; 3) Department of Psychiatry & Behavioral Neuroscience, University of Chicago; 4) Department of Medicine, University of Illinois at Chicago.

Nearly 1/5 Americans smoke cigarettes and 443,000 Americans die of smoking every year. Several pharmacotherapies substantially increase smoking cessation rates. However, treatment effects vary across ethnic groups, especially African Americans (AA). To determine if genetic background explains these differences between AA and European Americans, we studied the effects of naltrexone therapy on smoking cessation rates among AAs with ancestry at different levels of west African (WA) genetic ancestry. We used data from a previously published randomized, double blind trial of naltrexone vs. placebo for smoking cessation in 315 smokers from Chicago. Ninety-five participants were AA and provided a DNA sample. We genotyped 105 autosomal DNA ancestry informative markers using the SEQUENOM iPLEX MassArray platform. Individual ancestry was estimated for all subjects using STRUCTURE. We used logistic regression models to estimate the effects of naltrexone on 4-week quit rates for AAs with high and low WA ancestry. The models adjusted for baseline clinical and demographic characteristics and confidence intervals and p-values were estimated using the bootstrap. In the total pooled sample of AAs, 4-week quit rates were not statistically different between naltrexone (41%) and placebo (33%) groups (p=0.44). However the pooled analysis masks treatment effect heterogeneity by WA ancestry. The median WA ancestry among the 95 AAs in our sample was 80%, so we defined subjects above the median as high WA ancestry and those below median as low WA ancestry. Among those AAs with low WA ancestry, quit rates were higher in the naltrexone group than the placebo group (59% vs. 29%) so that the treatment effect for those AAs with low WA ancestry was 30 percentage points (95%CI -0.02 to 0.58, p=0.04). In contrast, for those AAs with high WA ancestry, quit rates were actually lower in the naltrexone group (25% vs. 59%) so that the treatment effect was negative (-41 percentage points (95%CI -0.41 to 0.13, p=0.3)). The difference in naltrexone treatment effects on quit rates across high and low WA ancestry groups was statistically significant (p=0.04). Naltrexone efficacy on quit rates varies across AA subjects with different levels of WA ancestry. Our data suggest naltrexone is effective for AAs with low WA ancestry and is not effective in AAs with high WA ancestry. These results suggest that genetic factors may partially explain racial differences in treatment effects.

743W Assessment of viability of human lymphocytes exposed to ionizing radiation and curcumin. B. Gonzalez1, C. Monterrubio1, H. Pimentel1, R. Silva1, L. Rodriguez1, C. Barba1, N. Perez1, B. Inda1, J. Vidal1, E. Flores1, H. Reyes1, L. Bobadilla2, A. Corona2,1. 1) Laboratorio de Citogenética, Genotoxicidad y Biomonitorio, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Unidad de Citogenética, Serv. Hematol. Oncol. Pediátr., Div. Pediatría, Hosp. Civil Dr. Juan I. Menchaca, Guadalajara, Jalisco, Mexico.

Introduction: Curcumin is a polyphenol with antioxidant, anti-inflammatory and antineoplastic properties. In vitro it inhibits tumoral cell proliferation but in contrapart there are reports of cytotoxicity and genotoxicity effects on healthy human lymphocytes. Ionizing radiation induces cell damage and genotoxicity. Cytotoxicity can be assessed on individual cells by trypan blue testing or flow cytometry or assessment on cells population under XTT assay. Statement of Purpose: To evaluate the effect of curcumin on human lymphocytes viability under the effect of ionizing radiation with three viability assessment methods. Methods Used: Peripheral blood lymphocytes from 5 healthy individuals where obtained and cultured for 72hrs, under the next conditions: 1) control group, 2) 25µM of curcumin for 48hrs, 3) 2gy radiation after 72hrs culture, and 4) curcumin for48hrs with radiation after 72hrs. XTT, trypan blue, CD45+ and 7AAD flow cytometry were done for cell viability assessment. Results: Viability percentages were as follows: XTT testing showed viability of 18.5 for the cells under condition 2, 106.66 for condition 3, 11.23 for condition 4 and 100 for control. Flow cytometry showed viability of 41.84 for condition 2, 60.74 for condition 3, 36.16 for condition 4, and 66.34 on control group. Under trypan blue: condition 2 had viability of 53.07, condition 3 of 63.05, condition 4 of 39.37, while control had a viability of 94.96 from TaqMan Genotyping Software and of 96.74 comparing group 2 and control (p=0.003), and group 3 and control (p<0.001), the same found under XTT testing (group 2 vs control with p=0.009; group 3 vs control with p<0.001). Trypance blue results analysis only found significant difference between group 3 and control (p=0.022). Conclusions: Our findings show that curcumin under 25µM, for viability is reduced when compared with the control group with 3 methods as shown before. The viability of the group exposed to 2gy radiation showed no significant difference to control, regard- less of the lymphocyte from repair damage. Flow cytometry and trypan blue did not showed statistically significant differences between them, however, XTT results where different, which is related that the later study evaluates groups of cells, while the first two assess individual cells. Sebastiá and cols. in 2012 reported a cytotoxic effect of curcumin on cultures, which is supported by our results.
Ankylosing spondylitis (AS) is a common, highly heritable, arthritis. 27 susceptibility loci have been identified in and outside the MHC. The HLA-B27 allele is the major genetic risk factor to AS and its role in disease aetiology remains elusive. Potential mechanisms include disorders of the antigen presentation function of HLA class I proteins, or on abnormal intracellular effects unique to the B27 protein variant. Prior studies have suggested that other HLA-B alleles and MHC genes are involved in AS-susceptibility. In this study we aim to better define the MHC associations and to identify functional and potentially causal variants to shed light on the mechanisms by which the B27 molecule confers risk to disease. We genotyped 7,264 MHC SNPs in 9,069 affected subjects and 13,578 controls of European descent using the Immunochip. Genotyping was followed by imputation of HLA class I and II alleles and amino acid positions of HLA proteins. Association to disease susceptibility was assessed by logistic regression correcting for population structure. Association was observed with SNPs in the HLA-B locus (P=10-320). Analysis of HLA-B alleles revealed other non-B27 alleles affecting susceptibility of moderate effect size (B40.01, OR=1.13; *40.02, OR=1.59; *51.01, OR=1.36; *07.02, OR=0.77; *57.01, OR=0.73). After controlling for the associated haplotypes in HLA-B we observed independent association signals with SNPs in the HLA-A locus and in HLA-DRB1. Analysis of polymorphic amino acid positions demonstrated that the most significant polymorphisms in the HLA-B and HLA-DRB1 loci were amino acid residues located in the binding pocket of these molecules. We have previously shown that AS is associated with ERAP1 polymorphisms only in HLA-B27 carriers. Additionally, we show that amongst B27 negative cases, ERAP1 variants are AS-associated in B40-positive but not -negative subjects, indicating that ERAP1 variants also interact with B40. This study has identified susceptibility alleles in HLA class I and II loci. The identification of multiple HLA-B alleles affecting susceptibility contrasts with the previous observation of an inter-individual variation in the rate of pain crisis in SCD patients is heritable. We aimed to verify if there were any genetic variants associated with SCD-related pain crisis within genes already known to be implicated in pain, based on the Pain Genes Database (LaCroix-Fralish, M.L., Pain, 2007). For our study, we had access to three different genotyping datasets, all originating from the Cooperative Study of Sickle Cell Disease: (i) 1,434 individuals were genotyped on the IBC array, a gene-centric array targeting ~2,100 genes related to heart, lung and blood diseases, (ii) 1,409 individuals were genotyped on the exome chip and (iii) 1,279 were genotyped on the Illumina 610-Quad array. 873 patients were genotyped on the three platforms. To maximize the number of samples included in each analysis, we analyzed the three datasets independently. We concentrated our analysis on sickle cell anemia patients with a minimum follow-up time of a year. Because the clinical definition of pain crisis is variable, we only analyzed patients who never experienced pain crisis (controls) and patients with pain crisis rates above the 90th percentile (cases). We used simulations to quantify how phenotype misclassification within the middle of the pain rate distribution can affect statistical power. SNPs with a minor allele frequency above 5% located less than 50kb from a gene implicated in pain were included in the common variants analysis. We performed a logistic regression using as covariates: fetal hemoglobin levels, hematocrit, sex, age and principal components. The significance threshold of $7 \times 10^{-8}$ was determined by calulating the overall number of independent signals tested in the three datasets. We also conducted a rare variants analysis with SKAT-O. This analysis included functional variants with a minor allele frequency lower than 5%. No gene showed significant association ($1 \times 10^{-8}$). Given our limited power, we can only conclude that there are no variants of strong phenotypic effect in genes previously implicated in pain that associate with pain crisis rate in SCD patients.
747W
Genetic Linkage Analysis and Candidate Gene Approach In Inbred Puerto Rican Families with Congenital Scoliosis. J.E. Baez1, J.C. Orengo1, E. Suarez2, C. Burgos3, V. Francheschi3, A.S. Corrêa1,2,3.
1) School of Public Health, Ponce School of Medicine and Health Sciences, Ponce 00717-0211, PR; 2) Department of Microbiology, Ponce School of Medicine and Health Sciences, Ponce, PR; 3) Department of Molecular Medicine, La Concępcion Hospital, San German, PR; 4) Clinical Research Center, San Jorge Children's Hospital, San Juan ,PR.

Scoliosis can be either idiopathic or secondary to congenital vertebral abnormalities. Idiopathic typically occurs in children and adolescents who are otherwise healthy. Congenital is due to anomalous development of the vertebrae (failure of formation and/or segmentation). Incidence is estimated in 1/1,000 to 1/2,000 but the true incidence remains unknown. Etiology is unknown. Age and Range Mean 6 years (range few days of life to 25 years).Genes of the Notch signaling pathway have been identified as responsible, they are involved in various biologic processes, including somatogenesis, and neurogenesis. We have found compound heterozygosity status in two families with 7 affected individuals with single mutations on the MESP2, Hes 7 and Dll1 genes. Homozygous linkage analysis in three other inbred families have proved evidence of linkage (LOD scores 1.75 to 2.5) in chromosomes 2 and 14 respectively. Evidence of genetic linkage to these chromosomes is new and may help identifying new candidate genes responsible for vertebral malformations and congenital scoliosis. Understanding and defining the natural history of specific mutation(s) and the developmental (molecular) mechanisms in vertebral patterning, will aid in the identification of protective factors for normal spinal development and toward the prevention of disfiguring congenital scoliosis.

748T
GENE POLYMORPHISMS AS RISK FACTOR FOR EARLY PRIMARY OSTEOARTHROSIS OF THE KNEE IN ASIAN INDIANS. S. poornima, k. subramanyam, q. Hasaan. Genetics & Molecular Medicine, Kamineni Hospitals, Hyderabad, India.

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

749F

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

750W
Role of the DIVERSIN gene in neural tube defects in humans. R. Allache1,2, P. De Marco1, E. Morello1, V. Capra3, Z. Kibar2. 1) Pathology and Cell Biology, University of Montreal, Montreal, Quebec, Canada; 2) Obstetrics and Gynecology, CHU Sainte Justine Research Center and University of Montréal, Montreal, Quebec, Canada; 3) Neurosurgery Department, G. Gaslini Institute, Genova, Italy.

DIVERSIN (DIV) is a core member of the planar cell polarity (PCP) pathway that controls the process of convergent extension (CE) during gastrulation and neural tube closure in vertebrates. It acts as a molecular switch where it activates the non-canonical Wnt/PCP pathway while simultaneously inhibiting the canonical Wnt/β-catenin pathway. The Wnt/PCP pathway has been strongly implicated in the pathogenesis of neural tube defects (NTDs) in animal models and human cohorts. In this study, we analyzed the role of DIV in these malformations by re-sequencing analysis of its open reading frame and exon-intron junctions in a cohort of 450 unrelated NTD patients. We identified 4 rare heterozygous and 1 homozygous missense mutations predicted those mutations to be functionally deleterious. Functional validation of these mutations on both Wnt/PCP and Wnt/β-catenin pathways using gene reporter assays is currently underway. Our study demonstrates that DIV could act as predisposing genetic factors to NTDs in a subset of patients and further expands our knowledge on the role of PCP genes in the pathogenesis of these malformations.
Epistatic interactions between SNPs in PHF11 and IFNG genes increase risk to allergic phenotypes. AK. Andiappan1,2, KJ. PUAN1, WS. Yeo1, BT. Lee1, R. Meichtry2, M. Poidinger1, D.Y. Wang1, O. Ratzschke1,2, C. Chew1,2. 1) Singapore Immunology Network, Singapore, Singapore; 2) Department of Biological Sciences, National University of Singapore; 3) Department of Otolaryngology, National University of Singapore.

Allergic diseases, such as asthma and allergic rhinitis, are complex diseases resulting from interactions between genetic and environmental factors. According to Allergic Rhinitis and its Impact on Asthma (ARIA), asthma and AR affect nearly 300 million and 500 million of the world population respectively. Genetic variants in PHF11 and IFNG have been associated with asthma and other allergic phenotypes in multiple populations. Here we report a case control study to evaluate the association of PHF11 and IFNG SNPs to allergic rhinitis and asthma either through individual SNPs or through SNP-SNP epistasis using a population of 2880 ethnic Chinese in Singapore with 2163 having atopy, 795 with AR and 718 having asthma. A total of 8 tagSNPs from PHF11 and 3 from IFNG gene were selected based on linkage disequilibrium. We used PLINK software to evaluate single SNP and SNP-SNP interaction in association to phenotypes. Our association results show that none of the SNPs were significantly associated (P<0.05) individually to any of the allergic phenotypes. In contrast however we identified SNP pairs interacting to increase risk significantly to atopy, asthma and AR (table attached). We also used Synergy Factor (SF) to evaluate the strength of interaction, which allows assessment of binary interactions in case-control studies considering both size and significance. The results validated that the SNP pair rs9526569 (PHF11) and rs2069728 (IFNG) increased risk with SF= 1.76 (P=0.01) for atopy, SF= 1.92 (P=0.007) for AR and SF= 2.22 (P=0.003) for asthma. The most important result was that the role of FOXO3A in longevity by 1) replicating the known associations, 2) trying to identify new variants of this gene associated with longevity, and 3) attempting to build a network model of the relationships among FOXO3A alleles, longevity and some of its sub-phenotypes in the Long Life Family Study (LLFS). The association analysis (N=4656) was performed using Cox proportional hazard models for age at death, censored at the last age at contact for living subjects, adjusting for sex and the top 10 principal components. Correlation within-family was modeled using random effects with variance covariance matrix proportional to the kinship matrix. Gene-wide significance threshold was determined by applying Bonferroni correction to the number of independent loci (M=20) in FOXO3A, estimated by the simpleM method (Gao et al. 2010). In the LLFS, 6 out of 11 known variants replicated (p<0.05), and 3 were associated with increased risk (p<0.0027, HR=0.83), but no novel variants were identified in 113 tested SNPs. Next we tried to dissect the significant associations between independent FOXO3A alleles and lifespan by building a network model (N=3136) that included all SNPs and at enrollment and at follow-up. The network included 8 SNPs (two from FOXO3A, six from CAD and ESRD). The analysis suggests that different alleles may affect exceptional longevity by 1) replicating the known associations, 2) trying to identify new variants of this gene associated with longevity, and 3) attempting to build a network model of the relationships among FOXO3A alleles, longevity and some of its sub-phenotypes in the Long Life Family Study (LLFS). The association analysis (N=4656) was performed using Cox proportional hazard models for age at death, censored at the last age at contact for living subjects, adjusting for sex and the top 10 principal components. Correlation within-family was modeled using random effects with variance covariance matrix proportional to the kinship matrix. Gene-wide significance threshold was determined by applying Bonferroni correction to the number of independent loci (M=20) in FOXO3A, estimated by the simpleM method (Gao et al. 2010). In the LLFS, 6 out of 11 known variants replicated (p<0.05), and 3 were associated with increased risk (p<0.0027, HR=0.83), but no novel variants were identified in 113 tested SNPs. Next we tried to dissect the significant associations between independent FOXO3A alleles and lifespan by building a network model (N=3136) that included all SNPs and at enrollment and at follow-up. The network included 8 SNPs (two from FOXO3A, six from CAD and ESRD). The analysis suggests that different alleles may affect exceptional longevity by 1) replicating the known associations, 2) trying to identify new variants of this gene associated with longevity, and 3) attempting to build a network model of the relationships among FOXO3A alleles, longevity and some of its sub-phenotypes in the Long Life Family Study (LLFS).

Genetic variation plays a major role in determining the susceptibility to the obese environment. Significant SNPs associations with BMI, FTO rs9939609-FTO with body mass index (BMI) and with the risk for obesity have been suggested in homozygotes AA. The central role of FTO might be through an effect on cerebrocortical insulin sensitivity; as homozygous have a reduced insulin secretion. Moreover, evidence has been found between 5-HTTPLPR polymorphism of the serotonin transporter SLC6A4 and being overweight, hypothesizing that it regulates behavioral and metabolic responses associated with the development of obesity through hunger and satiety. Body obesity is associated with oxidative damage and that PON1 has been found in the interstitial space between unilocular fat cells of the adipose tissue; some PON1 polymorphisms have been associated with obesity. In this study, we evaluated the association of these three polymorphisms with the risk for obesity in children from Yucatan, Mexico; where child obesity is the first cause of morbidity. We included 76 obese children, and 155 non-obese healthy children under a case-control association study. Genotype and allele frequencies between cases and controls were compared using SNPassist software. Genotype and allele frequencies were distributed according to Hardy-Weinberg expectations (p>0.05) in cases and controls, except for rs9939609-FTO in controls (p=0.007). Risk alleles frequencies in the studied population were: 16.45% for the SNP rs9939609-FTO; and 36.70%; for both the SNP rs705379-PON1 and SLC6A4 (5-HTTPLPR). Significant differences were found in the heterozygous AT genotype of the SNP rs9939609-FTO (p=0.11) between cases and controls, suggesting that the heterozygous AT genotype of FTO is a genetic risk factor an associated with child obesity in the population of Yucatan. TB and PTST- are associated with both PTST- and TB. Among the 583 subjects, 299 households included in this analysis were enrolled between 1995-1999 and 2002-2008. Analysis was conducted using general estimation equations, clustering the data by family. For each outcome, TB and PTST-, both the dominant and the additive model were run, with sex and HIV as covariables. We found SNPs in IL12RB2 and NOD2 to be associated with both PTST- and TB. 3 SNPs in IL10, 3 in SLC6A3, and 2 each in NOD1 and NOD2 were found to be associated with PTST- with IL12A, IL12B, and TLR pathway genes, associated with TB. Among the most significant SNPs associated with TB were IL12RB2 rs2307147 (3.5e-7), and TLR2 rs1816702 (9.4e-7). For PTST- the most significant were TOLLIP rs7543942 (p=0.012), SLC6A3 rs409588 (p=0.014), and IL12RB1 rs17852635 (0.019). Though these were not significant after correcting for multiple tests, some of these regions identified are under-studied, which provides an impetus to focus future research in these regions.

576W Mannose-binding lectin (MBL2) gene variations and malaria risk in Indian populations. A.N. Jha1, P. Sundaravadivel1, L. Singh, T.P. Vela van2, K. Thangaraj1, 1) Evolutionary and Medical Genetics, Centre for Cellular and Molecular Biology, Hyderabad, AP, India; 2) Institute of Tropical Medicine, University of Tübingen, Tübingen, Germany.

The pathogenesis of malaria and its severity depends on complex interplay of host genetic make-up, the parasite virulence and transmission dynamics and the host immune responses. MBL is a pattern recognition receptor of the innate immune, which, recognize and react to specific repertoire of carbohydrates on the surface of invading organisms and plays an important role in the course of infectious diseases. In this study, we aim to investigate the association of MBL2 variants with P. falciparum malaria infection in Indian populations. We re-sequenced the complete MIF gene along with 1 kb each of 5’ and 3’ region in 434 individuals from malaria endemic regions of the Orissa and Chhattisgarh states of India. The subjects comprised of 176 cases of severe malaria, 101 of mild malaria, and 157 ethnically matched asymptomatic controls. Since the entire Indian subcontinent represents malaria endemic region, we extended our study to 830 individuals from 32 socially, linguistically and geographically diverse Indian populations. The MBL2*221C (X) allelic variant is associated with increased risk (mild malaria OR:1.9, Pcorr=0.0036; severe malaria OR:1.6, Pcorr=0.02). The functional variant MBL2*B increases the risk (mild malaria OR:2.1, Pcorr=0.036; mild vs. severe malaria OR:2.5, Pcorr=0.039) and MBL2*C increases the risk towards severity (mild vs. severe malaria OR:5.4, Pcorr=0.045). The exon1 (MBL2*D/B/C) increases the risk towards severe malaria (OR:3.4, Pcorr=0.000045). The lower MBL secretor haplotypes were observed more in severe malaria compared to asymptomatic (OR:2.0, Pcorr=0.00002). The MBL2*LYPA significant differences in Pcorr for MBL2*LYPA increases the malaria risk. Further, the frequency of low MBL2 haplotype varies significantly among Indian populations (0 - 62.5%) and MBL2 variant MBL2*B is the prevalent form of structural variant in most of the Indian populations (3%) and the MBL2*C variant is significantly different in the studied population. We demonstrated the contribution of the heterozygous AT genotype of the SNP rs9939609-FTO and of the short allele S of the 5-HTTPLPR polymorphism, to the genetic risk for child obesity in the population of Yucatan, Mexico.
Cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene and Graves disease: Case-control study. Y. Lee1,2,3, C. Huang1,4, W. Ting1,4, F. Lo5, C. Lin6, Y. Wu7, M. Chien8, C. Wang9, W. Chen1, W. Lin1, S. Chang3, C. Lin1. 1) Dept Pediatrics & Med Res, Mackay Memorial Hosp, New Taipei City, Taiwan; 2) Department of Biomedical Sciences, Mackay Medical College; 3) Department of School of Medicine, Taipei Medical University; 4) Department of Pediatrics, Mackay Medicine, Nursing and Management College; 5) Department of Pediatrics, Chang Gung Memorial Hospital; 6) Department of Pediatrics, Mackay Memorial Hospital HsinChu Branch; 7) Department of Pediatrics, Changhua Christian Hospital; 8) Department of Endocrinology and Metabolism, Mackay Memorial Hospital, Taiwan, Taiwan.

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

758F


Recent studies have provided evidence that high density lipoproteins (HDL) regulate plasma glucose level by affecting insulin action, non-insulin mediated glucose uptake and insulin secretion and consequently affect the risk for type 2 diabetes (T2D). A recent longitudinal study reported that associations had on average 0.14% increase in HDL at baseline and this effect was stronger in women. In Pima Indians, high HDL cholesterol levels had a protective effect against T2D only in women. The objective of the present study was to analyze variants associated with lipid levels in a prior GWAS of 1100 Pima Indians to determine whether any of these variants increase risk for T2D. The strongest finding to date is with a SNP (rs6499863) in the cholesterol ester transfer protein gene (CETP). CETP is a known genetic and functional locus for HDL levels with higher CETP levels inversely related to HDL levels. Genotyping of this SNP in 3519 full heritage Pima Indian samples informative for T2D and lipid levels confirmed a strong association of the T allele with lower HDL levels (p = 5.6×10^-11) and identified an association with higher risk for T2D (p = 4.6×10^-4). The evidence for T2D association coming only from female subjects (N=1949; p=2.7×10^-6; 1.44[1.21-1.71]). We replicated the female-specific association with T2D by genotyping a non-overlapping sample of 3901 (2111 female) mixed heritage American Indians (in females, p = 0.036; 1.29[1.01-1.55]). The strongest evidence for a T2D associated allele came from combination of both data sets. In a separate association of HDL levels independently replicated in the additional samples, where the allele predicting lower HDL levels increased risk for T2D, there was no evidence for sex interaction in the association of genotypes with HDL. These observations show with prior finding that lower HDL concentration was a stronger predictor of future T2D in women as compared to men. rs6499863 maps to the S’ near gene of CETP in close proximity to a distant promoter. Our future studies will involve dense mapping in and around this region and studying functional variants in CETP for their role in T2D.

759W

Identification of genes that contribute to diabetic nephropathy in type 2 diabetes through family-based association testing of genetic variants detected by targeted exome sequencing. M.G. Pezzolesi1,2, J. Jeong1, A.M. Smiles1, J. Skupien1,2, J.C. Mychaleckyj1, S.S. Rich3, J.H. Warram1, A.S. Krolewski1,2. 1) Section on Genetics and Epidemiology, Joslin Diabetes Center, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA.

Genome-wide linkage analysis in extended pedigrees from the Joslin Study of Genetics of Nephropathy in Type 2 Diabetes Family Collection previously identified four genomic regions with evidence for linkage to albumin excretion levels (chromosomes 5q, 7q, 21p, and 22q) and two genomic regions linked to variation in renal function (chromosomes 2q and 7p). To identify the susceptibility genes that contribute to these diabetic nephropathy (DN) sub-phenotypes, we performed targeted exome sequencing of all protein-coding genes across these loci in this collection. One-hundred-twenty-six DN cases from 42 families with an excess of renal disease (~3 cases per family) were selected for resequencing of the coding region of 361 genes across the six linkage regions using a custom target enrichment library followed by next-generation sequencing. The resulting data achieved greater than 70X on-target read depth and exceeded 20-fold sequence coverage across 70% of the target regions. Multi-sample variant calling using GATK identified a total of 2,442 non-reference variants, including 966 potentially damaging variants. Among these, 14 rare functional variants were significantly over-represented in T2D DN cases relative to the NHLBI’s Exome Sequencing Project reference panel. To improve our ability to detect associations at these loci, we are currently expanding our sequencing efforts to all diabetic family members with proteinuria or end-stage renal disease (P = 0.016). The strength of this association improved further when all relatives were analyzed together (P = 3.82×10^-5). These preliminary data suggest that variants in this gene may contribute to the linkage peak identified at this locus. We anticipate that complete sequencing of the genes across the six linkage regions identified in the Joslin Study of Genetics of Nephropathy in Type 2 Diabetes Family Collection in all members of these families will allow us to establish the genes that contribute to variation in urinary albumin excretion and renal function at these loci.
760T
Imputation of Turkish Population Genotypes Using Immunochip Data and 1000 Genomes Reference Reveals Behçet’s Disease Association of SNPs in the EGR2 Locus.

E.F. Remmers1, M. Takeuchi2, D. Ustek3, N. Mizuki3, I. Tugay-Tutkun4, E. Seyahi5, Y. yzygzan5, A. Gül6, D.L. Kastner7. 1) Inflammatory Disease Sect, NHGRI/Med Genetics Branch, Bethesda, MD; 2) Dept. of Genetics, Institute of Experimental Medicine, Istanbul University, Istanbul; 3) Dept. Ophthalmology and Visual Science, Yokohama City University, Yokohama; 4) Istanbul Faculty of Medicine, Istanbul University, Istanbul; 5) Cerrahpaşa Faculty of Medicine, Dept., Istanbul University, Istanbul.

Behçet’s disease is a genetically complex multisystem inflammatory disease with common occurrence of orogenital ulcers, uveitis, and skin lesions. It can also involve inflammation of the gastrointestinal, pulmonary, musculoskeletal, cardiovascular and neurological systems. The disease is common among modern-day populations who live along the ancient silk trade routes and is a leading cause of blindness in these countries. We recently completed a genome-wide association study of Behçet’s disease in the Turkish population with a discovery collection of 1209 cases and 1278 controls genotyped for 311,459 SNPs for which 798,465 autosomal SNPs were imputed using a reference panel of 96 Turkish samples genotyped for this purpose. With independent replication data and meta-analysis the study identified disease-associated SNPs that exceeded genome-wide significance at genomic regions encompassing the major histocompatibility complex, IL10, IL23R, CCR1, STAT4, and the natural killer complex gene cluster on chromosome 12p13.2. In addition, variants in ERAP1 acted recessively and predominantly exerted their effects in individuals with the disease-associated HLA-B*51 type. We surmised that even with imputation, the 311K genotyped discovery SNPs do not fully inform the variation present in the population and predicted that additional loci could be identified by denser genotyping and by imputation using 1000 Genomes haplotypes as a reference. We genotyped 2014 cases and 1826 controls from Turkey using the Illumina Immunochip, which provides dense coverage for 186 immunologically relevant genes, and evaluated the loci for disease association. We found a chromosome 10 region encompassing ZNF365, ADO, and EGR2 contained many SNPs with near genome-wide significance. We therefore used IMPUTE2 to impute additional SNP genotypes with the 1000 Genomes reference haplotypes. The implication revealed 2 disease-associated SNPs with P < 5 x 10^-8. These SNPs clustered 5’ of the EGR2 gene (early growth response 2), which encodes a transcription factor, E3 SUMO-protein ligase. Defects in this gene are responsible for several Charcot-Marie-Tooth disease types and variants 5’ of the gene are associated with increased risk of Ewing sarcoma. Recent work has shown roles of EGR2 and EGR3 in controlling inflammation and in promoting T and B cell antigen receptor signaling, suggesting a contribution of EGR2 variants to the hyperinflammatory state of Behçet’s disease.

761F
Interrogating the PBX–WNT–TP63 pathway in human nonsyndromic cleft lip/palate.

R. Silva1,2,3, A. Letra1,2,3, L. Maini2, J.B. Mulliken3, S. Stiller4, S.H. Blanton5, J.T. Hochberg3. 1) Department of Endodontics, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 2) Center for Craniofacial Research, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 3) Department of Pediatric Dentistry, Pediatric Research Center, University of Texas Health Science Center Medical School at Houston, Houston, TX; 4) Children’s Hospital of Boston, Boston, MA; 5) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

Nonsyndromic cleft lip/palate (NSCL/P) is the most common human craniofacial birth anomaly and results from defects during morphogenesis of craniofacial structures. Mutations in genes involved in the morphogenesis and patterning of the craniofacial structures, such as the Wnt and Bmp gene families, irf6, and p63, lead to CL/P in mice and humans. Recently, we identified regulatory modules controlling limb morphogenesis. It is a leading cause of blindness in these countries. We recently completed a genome-wide association study of Behçet’s disease in the Turkish population with a discovery collection of 1209 cases and 1278 controls genotyped for 311,459 SNPs for which 798,465 autosomal SNPs were imputed using a reference panel of 96 Turkish samples genotyped for this purpose. With independent replication data and meta-analysis the study identified disease-associated SNPs that exceeded genome-wide significance at genomic regions encompassing the major histocompatibility complex, IL10, IL23R, CCR1, STAT4, and the natural killer complex gene cluster on chromosome 12p13.2. In addition, variants in ERAP1 acted recessively and predominantly exerted their effects in individuals with the disease-associated HLA-B*51 type. We surmised that even with imputation, the 311K genotyped discovery SNPs do not fully inform the variation present in the population and predicted that additional loci could be identified by denser genotyping and by imputation using 1000 Genomes haplotypes as a reference. We genotyped 2014 cases and 1826 controls from Turkey using the Illumina Immunochip, which provides dense coverage for 186 immunologically relevant genes, and evaluated the loci for disease association. We found a chromosome 10 region encompassing ZNF365, ADO, and EGR2 contained many SNPs with near genome-wide significance. We therefore used IMPUTE2 to impute additional SNP genotypes with the 1000 Genomes reference haplotypes. The implication revealed 2 disease-associated SNPs with P < 5 x 10^-8. These SNPs clustered 5’ of the EGR2 gene (early growth response 2), which encodes a transcription factor, E3 SUMO-protein ligase. Defects in this gene are responsible for several Charcot-Marie-Tooth disease types and variants 5’ of the gene are associated with increased risk of Ewing sarcoma. Recent work has shown roles of EGR2 and EGR3 in controlling inflammation and in promoting T and B cell antigen receptor signaling, suggesting a contribution of EGR2 variants to the hyperinflammatory state of Behçet’s disease.

762W
Genetic heterogeneity of midline facial defects with ocular hypertelorism.

M. Simion1, E.L. Freitas1,2, I. Lopes-Cendes1, V.L. Gil-da-Silva-Lopes1. 1) Department of Medical Genetics, FCM/UNICAMP, Campinas, Brazil; 2) Department of Genetics and Evolutionary Biology, Bioscience Institute, University of São Paulo (USP), São Paulo, Brazil.

Midline Facial Defects with Ocular Hypertelorism (MFDH) is a rare and heterogeneous association of features, with a wide spectrum of abnormalities. Most common ones are ocular hypertelorism, broadening of the nasal root and orofacial clefting, which suggest the involvement of genes related to regulation of embryological development of the frontonasal process. Structural and functional anomalies of the central nervous system (CNS) are also frequently in MFDH. Regarding the etiology of this condition, mutations in ALX1, ALX3 and ALX4 genes were described in some families with frontonasal dysplasia, the most recognizable form of MDFH. SHH, FGF8 and PAX3 genes participate of pattern and growth controlling during developmental facial primordia and central nervous system processes. Mutations in these genes are associated with other craniofacial disorders. Also, few cases of MFDH with chromosomal abnormalities have been reported. We performed array-CGH and investigated mutations in SHH, FGF8 and PAX3 genes by direct sequencing and in 14 patients with MFDH. Recognized monogenic disorders and well-known syndromic conditions were excluded. Array-CGH detected chromosomal abnormalities in almost 15% of sample: a deletion of 6.2 Mb at 2q36.1-q36.3; a deletion at 9p24.1-9p24.3 (6.36 Mb) and a duplication at 20p13 (14.85 Mb). In first case, PAX3 gene was located between the deleted region. Despite of several single nucleotide polymorphisms (SNPs), no pathogenic mutations were found, which could be related to the size of sample. However, these results reinforce the heterogeneity of MFDH and justified the indication of array-CGH in routine care and genetic investigation of these patients.

Background: A number of novel candidate genes have been identified in recent genome-wide association studies of suicide attempt in bipolar disorder (Willour et al., 2011; Perris et al., 2010). However, the historical candidate genes, including NTRK2, HTR1A, and HTR1B, did not appear to be among the top findings in these studies. Perhaps a combination of rare and common variants may contribute to the predisposition to suicidal behaviour. Method: We analyzed variant data from high-throughput DNA sequencing of 202 genes (Nelson et al., 2012) in our bipolar disorder cases of self-reported and genetically ascertained European ancestry. We analyzed the phenotype of suicide attempt as well as suicide severity score (from the Schedule for Clinical Assessment in Neuropsychiatry SCAN: 0=non-suicidal; 1=suicide plan/ideation; 2=suicide attempt without serious harm; 3=suicide attempt with serious harm; 4=suicide attempt designed to end life; N=227).

We conducted preliminary analysis using PLINK, with history of alcohol use disorder as well as sex and age included as covariates. A total of 3199 DNA variants across the 202 gene regions were analyzed. Results: Among the findings, we found a number of DNA variants in NTRK2 (rs41312188, rs2289656; p<0.05) and HTR1A (rs34118835; p<0.05) to be nominally associated with suicide severity scores. These regions are also nominally significant for the analysis of lifetime history of suicide attempt. Conclusions: We conducted a high-throughput targeted sequence analysis and validation of our initial results in other bipolar disorder samples. We will further explore the interaction of these gene variants with history of alcohol abuse/dependence.


1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Dept of Dermatology, University of Michigan, Ann Arbor, MI; 3) Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; 4) Toronto Western Hospital, Toronto, ON; 5) Department of Genetics, Washington University at St. Louis, St. Louis, MO; 6) School of Medicine, Dermatology, University of Utah, Salt Lake City, UT; 7) Memorial University, St. John’s, NF, Canada.

Psoriasis is a chronic, immune-mediated inflammatory disease of the skin with prevalence ranging from 0.2 to 2% in different populations. To date, genome-wide association studies have identified 36 psoriasis susceptibility loci in European populations, with several showing strong evidence for association at novel variants, indicating both genetic and allelic heterogeneity. In most cases, the precise causal variant at each locus are yet to be determined. With the goal of identifying these causal variants, we selected 100 candidate regions based on their association in the genome-wide association study analysis and containing 769 genes, for deep, targeted resequencing. In each region, we sequenced all exons of transcription units within 250 Kb of the strongest previous association signal; except in the MHC, where we targeted five specific candidate genes comprising 137 Kb of sequence. In addition, we used a whole-region sequencing strategy that included non-coding intergenic and intronic sequence for ten regions with the most significant association signals, and a full transcription unit plus flanking sequence approach for four intronic sequence for ten regions with the most significant association signals. In total, we typed a set of 52 MALD SNPs on chromosome 1 to re-assess suspected UL loci on this chromosome as well as 164 randomly selected MALD SNPs on chromosomes 15, 16, 18, 19 and 21. Analysis. The ADIMXRP program was used in case-only models to compare ancestry at test SNPs with genome-wide ancestry and statistical significance was assessed using standard normal Z statistics, with a threshold of Z >4.0 considered statistically significant. Results. In models with no stratification for body mass index (BMI), a phenotype associated with UL albeit in a complex relationship, SNP loci at several chromosomes (1, 2, 3, 4, 5, 11, 12 and 15) reached statistical significance (Z >4.0; p=5.5E-05-3.9E-15), with those on chromosomes 1p11, 2q37 and 12q24 being the most significant (Z >5.0). After stratification by BMI, only chromosomes 1q42, 2q37 and 15q24 were found to have strong signals across all the BMI-categorizes (BMI<25, 25-30 and >30), with the association peaking at rs756784 on 1q42 (Z=4.0-4.7; p=7.6E-05-1.4E-08), at rs256552, rs256550 and rs12479375 on 2q37 (Z=4.8; p=2.2E-02-1.7E-06), and at rs1562250 and rs8030499 on 15q24 (Z=2.4-3.2; p=1.8E-02-3.6E-03). Our data show that only rs756784 on 1q42 remains significant across the BMI categories after correction for multiple testing. Conclusion. While this study replicated previous findings, only the new candidate 1q42 remained statistically significant after correction for multiple testing. Co-localization of UL-associated phenotypes such as the age at menopause and obesity-related traits on 1q42 further supports a candidate status for this region.


Background. The genetic basis of susceptibility to uterine leiomyomata (UL or fibroids), a condition with an estimated cumulative incidence of more than 80% among African American women of age 50, remains largely unexplained. The higher frequency of UL among African Americans when compared to European American women prompted us to adopt admixture mapping to evaluate candidate chromosomal regions for UL. Study design. We genotyped a total of 1,583 MALD (mapping by admixture linkage disequilibri- um) markers, a set of 340 SNPs, in a sample of African American women obtained by ultrasonography in the NIH EUtens fibroid study. We selected 1,167 MALD SNPs overlapping candidate chromosomal regions recently identified by other groups through linkage scan (2q37, 3p21, 5p13, 10p11, 11p15, 12q14 and 17q25) and admixture mapping (2q37, 4p16.1, 10q26). We also geno- typed an additional set of 252 MALD SNPs on chromosome 1 to re-assess suspected UL loci on this chromosome as well as 164 randomly selected MALD SNPs on chromosomes 15, 16, 18, 19 and 21. Analysis. The ADIMXRP program was used in case-only models to compare ancestry at test SNPs with genome-wide ancestry and statistical significance was assessed using standard normal Z statistics, with a threshold of Z >4.0 considered statistically significant. Results. In models with no stratification for body mass index (BMI), a phenotype associated with UL albeit in a complex relationship, SNP loci at several chromosomes (1, 2, 3, 4, 5, 11, 12 and 15) reached statistical significance (Z >4.0; p=5.5E-05-3.9E-15), with those on chromosomes 1p11, 2q37 and 12q24 being the most significant (Z >5.0). After stratification by BMI, only chromosomes 1q42, 2q37 and 15q24 were found to have strong signals across all the BMI-categorizes (BMI<25, 25-30 and >30), with the association peaking at rs756784 on 1q42 (Z=4.0-4.7; p=7.6E-05-1.4E-08), at rs256552, rs256550 and rs12479375 on 2q37 (Z=4.8; p=2.2E-02-1.7E-06), and at rs1562250 and rs8030499 on 15q24 (Z=2.4-3.2; p=1.8E-02-3.6E-03). Our data show that only rs756784 on 1q42 remains significant across the BMI categories after correction for multiple testing. Conclusion. While this study replicated previous findings, only the new candidate 1q42 remained statistically significant after correction for multiple testing. Co-localization of UL-associated phenotypes such as the age at menopause and obesity-related traits on 1q42 further supports a candidate status for this region.
Discovery and fine-mapping of BMI loci using Metaobochip: a trans-ethnic meta-analysis from the Population Architecture using Genomics and Epidemiology (PAGE) Study. 1) Public Health Science, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Preventive Medicine, Keck School of Medicine / Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 3) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 4) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Department of Genetics, Rutgers University, Piscataway, NJ; 6) Department of Biostatistics, University of Washington, Seattle, WA; 7) The Human Genetics Center and Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX; 8) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 9) Department of Biostatistics, University of Alabama, Birmingham, AL; 10) Preventive Medicine and Epidemiology, Loyola University, Chicago, IL; 11) Division of Cardiology, Geneva University Hospital, Geneva, OH, Switzerland; 12) Department of Biostatistics, Washington University, St. Louis, MO; 13) Carolina Center for Genomic Sciences, University of North Carolina, Chapel Hill, NC; 14) Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC. 

Genome-wide association studies (GWAS) primarily performed in European-ancestry (EA) populations have identified numerous loci associated with body mass index (BMI). However, it is still unclear whether EA associations can be generalized to other ancestry groups. In addition, for the most part the putative functional variant(s) in these BMI loci remain under investigation. Different local Linkage Disequilibrium (LD) structure between ethnic groups can help narrow in or fine-map these BMI-related loci. Therefore, we used the Metaobochip to densely genotype and evaluate 21 BMI GWAS loci identified in EA studies in 29,151 African Americans (AA), 15,000 Asians, and 6,211 Hispanics from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. Our preliminary analysis showed that among the 21 loci, FTO was associated with BMI in all of the three ethnic groups; MC4R in AA and Hispanics; SEC16B, TMEM18, ETV5, GNPDA2, TFAP2B, and BDNF in AA (at nominal significance level: 0.05/average number of SNPs across 21 loci). Moreover, Metaobochip-wide trans-ethnic meta-analyses (fixed-effects models) revealed two novel BMI loci, COBBL1-GRB14 (rs10184304, p=2.2×10-7) and TCF7L2 (rs7903146, p=5.4×10-8) loci, that were significant when adjusting for the total number of SNPs tested across the chip (2.5×107). The analysis in AA showed that fine-mapping in AA is a powerful approach to narrow in on the underlying causal variants in BMI GWAS loci identified. We will expand upon this work by including additional Asian, Hispanic and Native American samples into this study (totaling ~20,000 Asian, 20,000 Hispanics and 500 Native Americans) providing improved power to study known loci and identify novel loci in a trans-ethnic meta-analysis.
769T

GTF2IRD1 is an epigenetic regulator involved in facial skin patterning that may underpin facial abnormalities of Williams-Beuren syndrome patients. C.P. Canales1, S. Corley2, P. Kaur3, I. Smyth4, M. Wilkins5, E.C. Hardeman1, S.J. Palmer1. 1) Cellular and Genetic Medicine Unit, University of New South Wales, Kensington, Sydney, New South Wales, Australia; 2) School of Biotechnology and Biomolecular Sciences, The New South Wales Systems Biology Initiative, University of Sydney, Sydney, Australia; 3) Epithelial Stem Cell Biology Laboratory, Peter MacCallum Cancer Centre, Melbourne, Australia; 4) Dept of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia.

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

770F

Probabilistic method with correction for imputation error improves fine-mapping resolution. B. Pasanovic1, G. Richaev2, N. Zaitlen3, G. Bhala4, R.扩充文本...
773F Polymorphisms of DNA leading to increased susceptibility for Type 2 Diabetes in South Asians: A systematic review and meta-analysis. Z.N. Sohani1,2,3, W.Q. Deng1,2, S.S. Anand1,2,3. 1) Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; 2) Population Genomics Program, McMaster University, Hamilton, Ontario, Canada; 3) Population Health Research Institute, Hamilton General Hospital, Hamilton, Ontario, Canada. 

Background: Type 2 diabetes (T2D) is a metabolic disorder resulting from interplay between pancreatic β cell function and insulin resistance. People of South Asian ancestry are up to four times more likely to develop T2D than white Europeans and are forecasted to carry 25% of the world’s burden by the year 2030. Single Nucleotide Polymorphisms (SNPs), which explain the most genetic variation in humans, may contribute to the increased susceptibility for T2D in South Asians. However, risk from genetic variants has not been fully quantified for this ethnic group. The purpose of this study was to i) ascertain risk estimates from T2D associated polymorphisms in South Asians, and ii) compare effect sizes, risk alleles, and risk allele frequencies (RAFs) from this meta-analysis with genome wide estimates in white Europeans. Methods: A MEDLINE, EMBASE, and CINHAL search identified 3498 studies. An additional 17 studies were identified by hand-search and consultation with experts. We included 88 case control studies, which met our inclusion criteria. Data were extracted independently by two reviewers. SNPs were meta-analyzed using a random effects model and weighted by inverse variance. Odds ratios were determined for the risk allele. Results: After considering the linkage disequilibrium structure, 264 SNPs were considered from all included reports comprised of 65317 cases and 86511 controls. From these, 71 SNPs were conducive to a meta-analysis. Significant association with T2D was found in 22 SNPs from 17 nuclear genes (ADCY5, CDKAL1, CDKN2A/2B, ENPP1, FTO, GLIS3, HHEX, IDE, IGF2BP2, ITLN1, KCNJ11, PGC1α, PPARY, SLAMF1, SLC30A8, TCF7L2, and UCP2) and 2 SNPs from mitochondrial DNA (mtDNA). The majority of SNPs from nuclear genes increased odds of T2D by between 15-35% per risk allele. Risk alleles from mtDNA SNPs increased odds of T2D by 90%. Interestingly, effect sizes in South Asians were similar to those reported in white Europeans. However, we observed variation in risk allele frequencies in South Asians and common allele in Europeans. Conclusions: This is the first meta-analysis to investigate polymorphisms associated with T2D in South Asians. We demonstrate that similar effect sizes for T2D SNPs are apparent across ethnicity, but there is variation in RAFs. Differences in risk allele for HHEX may be indicative of selection or a result of genetic drift. Further investigation is warranted.

774W Assessment of Cytochrome P450 Genetic Variability on Methadone Dose and Tolerance. H. Tsai1,3, S. Wang2, S. Liu1, H. Ho1,2, Y. Chang2,3, W.Q. Deng1,2, S.S. Anand1,2,3. 1) Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, NY, USA; 2) Department of Drug Safety, China Medical University Hospital, Taiwan; 3) Department of Psychiatry, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan.

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

775T Mapping expression quantitative traits in mesenchymal stem cell cultures derived from nonsyndromic cleft lip and palate patients’ orbicularis oris muscle. C. Masotti1, A.C. Nica2, S. Ferreira3, L. Bomílcara, D. Meyer3, D.Y. Sunaga1, D.F. Bueno1, N. Alonso2, D. Franco3, E. Dermitsakis4, M.R. Passos-Bueno1. 1) Department of Genetics, University of São Paulo, Univ de Sao Paulo, São Paulo, Brazil; 2) Department of Human Genetics and Development, University of Geneva, Switzerland; 3) Department of Plastic Surgery, Facult of Medicine, University of São Paulo, Brazil; 4) Department of Plastic Surgery, Hospital Clementino Braga Filho, Faculty of Medicine, Federal University of Rio de Janeiro, Brazil.

Mesenchymal stem cells derived from dental pulp of nonsyndromic cleft lip and palate (NS CL/P) patients exhibit a distinct expression signature at loci related to extracellular matrix modeling and epithelial-mesenchymal transition processes. Orbicularis oris muscle is compromised in NS CL/P patients and is an accessible source of mesenchymal stem cells. In this work we established mesenchymal stem cell cultures derived from orbicularis oris muscle (OOMMSC) of 43 NS CL/P patients and four controls in order to identify genetic variants associated with gene expression variation and disorder susceptibility. By correlating genome-wide expression and genotype microarrays, we mapped 119 expression quantitative trait loci (eQTLs) related to 18 genes (p<0.0001; FDR=14%). We did not observe significant enrichment of any GO term among genes for which we identified eQTLs. Thirty out of the 119 eQTLs are shared with other tissues or cell types, such as lymphoblastic cell lines, liver, brain cortex, T-cells, and fibroblasts. Twelve previously reported NS CL/P candidate genes have at least one OOMMSC eQTL located within their putative regulatory regions, but none of these genes are directly regulated by these eQTLs. Four eQTLs for SENP5 and CEP19 genes (p<0.05, FDR<5%) are in linkage with NS CL/P genome-wide associated SNPs in 3q29 region (r2>0.52; D’>0.9). The haplotype containing the at-risk variant (rs10489880) is associated with higher expression of SENP5 and lower expression of CEP9. SENP5 is a SUMO-conjugating enzyme which plays a critical role in the regulation of sumoylated protein levels in the cell, and sumoylation was demonstrated as an essential mechanism for palate development. Further case-control association study of OOMMSC eQTLs will elucidate the potential role of these variants in NS CL/P susceptibility. Funding: FAPESP/CNPq.
776F Variations in ORAI1 gene associated with Kawasaki disease. Y. Onouchi,1,2,4, R. Fukazawa2,4, K. Ozaki3, M. Terai2, H. Hamada3, T. Honda2, H. Suzuki2, T. Suenga3, T. Takeuchi3, K. Yasukawa2, R. Ebata2, K. Higashi4, T. Saji1,2,4, Y. Kimmotsu1,2,4, S. Takatoku2,4, K. Ouchi5, F. Kishi6, T. Yoshikawa7, T. Nagai1, K. Hamamoto7, Y. Sato7, J. Abe8, M. Seki8,17,12, T. Kobayashi8, Y. Nakamura8, A. Hatasa6, T. Tanaka9,12,3,1 Lab. for Cardiovascular Medicine, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 2) Dept. of Pediatrics, Nippon Medical School, Tokyo, Japan; 3) Dept. of Pediatrics, Tokyo Women’s Medical University, Yachiyo Medical Center, Chiba, Chiba, Japan; 4) Dept. of Pediatrics, Kawasaki Medical University, Wakayama, Wakayama, Japan; 5) Dept. of Pediatrics, Chiba University Graduate School of Medicine, Chiba, Chiba, Japan; 6) Dept. of Cardiology, Chiba Children’s Hospital, Chiba, Chiba, Japan; 7) Dept. of Pediatrics, Toho University School of Medicine, Tokyo, Japan; 8) Dept. of Pediatrics, Kawasaki Medical School, Kurashiki, Okayama, Japan; 9) Dept. of Molecular Genetis, Kawasaki Medical School, Kurashiki, Okayama, Japan; 10) Dept. of Pediatrics, Fujita Health University, Toyoake, Aichi, Japan; 11) Dept. of Pediatrics, Dokkyo Medical University Koshigaya Hospital, Koshigaya, Saitama, Japan; 12) Dept. of Occupational Therapy, International University of health and welfare, Fukuoka, Fukuoka, Japan; 13) Dept. of Pediatrics, Fuji Heavy Industry LTD, Health Insurance Society General Ohta Hospital, Ohta, Gunma, Japan; 14) Dept. of Allergy and Immunology, Division of Immunology, National Research institute of Child Health and Development, Tokyo, Japan; 15) Dept. of Cardiology, Gunma Children’s Medical Center, Shibukawa, Gunma, Japan; 16) Division of Clinical Pharmacology and Toxicology, The Hospital for Sick Children, Toronto, Ontario, Canada; 17) Div. of Nephrology, Department of Medicine, Maebashi, Gunma, Japan; 18) Lab. for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 19) Lab. for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 20) Lab. for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 21) Lab. for Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan; 22) Lab. for Cardiovascular Medicine, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 23) Dept. of Human Genetics and Disease Diversity, Tokyo Medical and Dental University, Tokyo, Japan; 24) Japan Kawasaki Disease Genome Consortium.

Kawasaki disease (KD; MIM611775) is a systemic vasculitis syndrome with unknown etiology which predominantly affects infants and children. Recent findings of susceptibility genes for KD have suggested up-regulation of Ca2+ entry located on 1q24 where positive linkage signal was seen in our previous sib pair study of KD, and conducted a genetic association study. By re-sequencing 23kb of ORAI1 region for 24 subjects including 12 KD cases and 12 controls, we identified 68 variants and then performed linkage disequilibrium analysis with 94 Japanese healthy individuals. After selecting 9 tagging SNPs which represent 37 variants with minor allele frequencies larger than 5%, we performed an association study using 730 KD cases and 1315 controls. Among the 9 SNPs examined, a non-synonymous SNP (rs3741596; p. S218G) showed a nominal association with KD (OR=1.92, 95%CI 1.02~1.40, P = 0.028). The same trend of association was observed in an independent case control panel (1586 KD cases and 3194 controls) and a sex specific analysis was also conducted. Meta-analysis (OR=1.20, 95%CI 1.08~1.33, P = 0.0098). Furthermore, we also found a rare 6 base-pair insertion polymorphism which cause elongation of ROAI1 protein was overrepresented in KD cases (rs7844824; OR = 3.91, 95%CI 1.30~11.80, P = 0.010). These data indicate altered ORAI1 function confers susceptibility of KD and further highlight importance of the Ca2+ entry pathway in the disease pathogenesis.


Antinuclear antibodies (ANA) are the serologic hallmarks of autoimmune and connective tissue diseases, Center for Integrative Medicine, University of Texas, Dallas, TX, 2) Division of Rheumatology, Peitian Japanese Cardiovascular Hospital, 500 University Drive, Hershey PA 17033, 3) Oklahoma Medical Research Foundation, 825 North East 13th Street, Oklahoma City, Oklahoma 73104, USA; 4) Service de Rhumatologie, SSS/IREC/RUMA, Cliniques Universitaires Saint-Luc & Institut Rubens, Brussels, Brussels, Belgium; 5) Feinstein Institute for Medical Research in Manchuset, New York, USA.

Systemic lupus erythematosus (SLE) is an autoimmune disease with a strong genetic component and dramatic gender difference in disease prevalence, including male controls were also used for replication to explore possible gender differences. Variants in or near PRP2, NAA10 and THEM1 were shown to be associated with SLE in Asian populations, surpassing genome-wide significance. In addition, two previously reported loci in European populations, IRAK1 (rs1059702), and MECP2 (rs2734647) were also replicated in our study. We also identified a variant upstream of L1CAM with suggestive association, which showed clear interaction with SNPs in NAA10. Independent contributions of these variants were tested using various methods and found a rare 6 base-pair insertion polymorphism which cause elongation of ROAI1 protein was overrepresented in KD cases (rs7844824; OR = 3.91, 95%CI 1.30~11.80, P = 0.010). These data indicate altered ORAI1 function confers susceptibility of KD and further highlight importance of the Ca2+ entry pathway in the disease pathogenesis.
1) Division of Biostatistics, Washington Univ in St Louis, St Louis, MO; 2) Department of Genetics, Washington Univ in St Louis, St Louis, MO; 3) Department of Medicine, Medical College of Wisconsin, Milwaukee, WI; 4) Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL; 5) Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL; 6) Department of Internal Medicine, University of Utah, Salt Lake City, Utah.

Background: Candidate variants from next generation sequencing (NGS) studies of complex diseases poses a great challenge because of the large number of rare variants, effects of which may vary in size and direction. Many existing methods 'collapse' clusters of rare variants to detect aggregate effects, and their power can be seriously compromised when the clusters contain only a few informative variants or when the effects are inconsistent.

On the other hand, set enrichment analysis, which has been widely used for microarray gene expression and pathway-based genome-wide association studies, remains powerful even when a large number of irrelevant variants exist in the set. Methods: A recently developed method called Variable Set Enrichment Analysis (VSEA) was applied to the exome sequencing data of 21 African American subjects from the HyperGEN (Hypertension Genetic Epidemiology Network) study, to prioritize genes associated with a left ventricular mass trait (LVMMT2). It evaluates the enrichment of association signals indiscriminately of rare or common variants within genes, based on SNP level statistics. We tested 3 SNP level statistics including the t-score, log likelihood ratio, and the regression coefficient in a linear model. Their performance was compared with that of a step-up collapsing method. Results: For chromosomes 17 and 18 are completed and reported below. Results: Among the 1192 genes that have data on chromosomes 17 and 18, VSEA test returns 52~68 genes with false discovery rates (FDR) q-value<0.05, of rare and common variants. While more research is needed to evaluate the variable set enrichment analysis empirically in presence of large number of irrelevant variants in the sample size, VSEA detected many known candidate genes for heart disease (p=0.042~0.046). Conclusion: The analyses demonstrate power of the solute-carrier family, and the Wnt family, were also identified, though IGFBP1(0.0051), ALOX15(0.021), and LPO(0.041). Genes from several of these families, such as the calcium channel, potassium channel, integrins, the solute-carrier family, and the Wnt family, were also identified, though the relationship of these genes with heart diseases is less clear. In contrast, the test using step-up collapsing only discovered two solute-carrier family genes (p=0.041). These results demonstrate the advantage of the variable set enrichment analysis empirically in presence of large number of rare and common variants. While more research is needed to evaluate the influence of SNP-level test statistics, the VSEA approach seems to provide a simple and effective way to prioritize genes in NGS studies.

A replication study for four keloid loci at 1q41, 3q22.3-23 and 15q21.3 was genotyped using Small Amplicon Genotyping method. In this study, the two from four loci identified by previous GWAS, rs873549 on chromosome 1 (odds ratio (OR) =4.8, 95% CI: 1.941-14.192) in recessive manner and rs8032158 located in NEDD4 on chromosome 15 (OR= 3.019, 95% CI: 1.712-5.308 ) in dominant manner were significantly associated in keloid compared with control. Especially, in rs8032158, the risk for severe type of the scar severity scale (OR= 2.563, 95% CI: 1.264-5.238) increased in comparison with mild type. Individuals with two risk alleles of the two loci showed higher risk of developing keloid than those with one risk allele alone. Our findings elucidated the significance of genetic variation at these 2 loci in keloid in the Japanese population.

Identification of potentially causative variants underlying triglyceride levels. S.M. Raj1, A. Coventry1, C. Ballantyne2, C.F. Sing2, R. Gibbs3, E. Boerwinkle4, A. Clark1, 1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI 48105; 4) Human Genetics Center, Health Sciences Center, University of Texas, Houston, TX 77030.

There is enormous incentive to identify the causative variants that underlie associations identified in GWA studies. Here we present an approach to optimize the allocation of sequencing resources for GWAS follow-up studies. The method specifies both informative individuals and the specific genomic regions to resequence, and it specifies an analytical approach to identify phenotypic effects. This approach was applied to a known GWAS signal for serum triglyceride levels in European-Americans from the Atherosclerosis Risk in Communities (ARIC) study. We also explored methods for querying GWAS genotype data to identify rare variants responsible for differences in triglyceride levels. We first carried out single-marker association tests to identify markers that show the strongest association with triglyceride levels. The shared haplotype structure in the region of these associations specified a 110 kb region on chromosome 7 that we sequenced (30-fold depth) in 502 individuals from the tails of the distribution of triglyceride levels. Analysis of these data revealed strong associations with TRIG levels at candidate sites that also showed strong prior biological relevance. We identified a 2 kb region in 8CL7B which is the target of many transcription-factor binding sites as a candidate for influencing triglyceride biology. The method used here lends stronger support for candidate regions to motivate biochemical or molecular-based validation of their role in disease etiology. Identification of potentially causative variants may therefore be useful in developing therapeutical applications.
783W
The contribution of regulatory variation to facial masculinity in humans. A.A. Zaidi, P. Claes, W. Yao, K. Daniels, M.D. Shriver. 1) Genetics, Pennsylvania State University, University Park, PA, United States; 2) University of Leuven, Leuven, Belgium; 3) Anthropology, Pennsylvania State University, University Park, PA, United States.

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

783F
Association of genetic variations in anti-inflammatory cytokine pathway genes in the outcome of tuberculosis. S. Aggarwal1,2, S. Ah, R. Chopra, A. Srivastava, R.N.K. Bamezai3. 1) Department of Biochemistry, All India Institute of Medical Sciences, Raipur, 492099 Chattisgarh, India; 2) National Centre of Applied Human Genetics, School of life Sciences, Jawaharlal Nehru University, New Delhi, 110067, India.

Tuberculosis (TB) constitutes the major cause of death due to infectious diseases. Cytokines play a major role in defense against Mycobacterium tuberculosis infection. Polymorphisms in the genes encoding various cytokines have been associated with tuberculosis susceptibility or resistance. This study aimed to investigate potential associations of forty SNPs in anti-inflammatory cytokines (IL10, TGFβ1, IL6, IL4 and IL13) and receptors (IL10RA, IL10RB, TGFβR1, TGFβR2, IL6R, IL4R, IL5RA, IL5RB and IL13RA1) with tuberculosis. Our study included 327 clinically categorized tuberculosis patients representing North Indian population and 1294 unrelated healthy individuals. Analysis of rs3024498 located in the 3'UTR region of IL10, was observed to be significantly associated with the risk towards tuberculosis (AA/AG+GG, OR=1.52, 95%CI=1.11-2.10, P=0.01). Other polymorphisms, rs7281762 of IL10RB, (GG/GA+AA, OR=0.7, 95%CI=0.55-0.89, P=0.004) provided protection and rs2228048 of TGFβR2 (TT/CT+CC, OR=2.7, 95%CI=1.2-6.02, P=0.02) risk towards the disease, respectively. Our results demonstrate that the polymorphisms in cytokine genes may be valuable markers to predict the protection or risk towards tuberculosis. In conclusion, this study provides an interesting cue to cumulative polygenic host component which regulates tuberculosis pathogenesis.

784T
Genetic associations with high-grade dysplasia and colorectal cancer in patients with colonic inflammatory bowel disease: preliminary results from ImmunoChip using a targeted analytic approach. J. Knight1,2, M. Waterman3,4, J.M. Stempak5, K. Krishnaprasad6, I. Cleynen7, L.P. Schumm1, S. Vermaier1, D.P. McSovem1, S.R. Brant1,3, G. Radford-Smith2, M.S. Silverberg1,2, International IBD Genetic Consortium. 1) Neuroscience Research, Centre of Addiction and Mental Health, Toronto, Ontario, Canada; 2) University of Toronto, Toronto, ON, Canada.; 3) Division of Gastroenterology, Mount Sinai Hospital, Toronto, Toronto, ON, Canada; 4) Department of Gastroenterology, Rambam Health Care Campus, B. Rappaport Faculty of Medicine, The Technion - Israel Institute of Technology, Haifa, Israel; 5) Inflammatory Bowel Diseases Research Group, Queensland Medical Research Institute, Brisbane, QLD, Australia; 6) Translational Research in Gastrointestinal Disorders, K.U.Leuven, Leuven, Belgium; 7) University of Chicago, Chicago, IL.; 8) Department of Gastroenterology, University hospitals, Leuven, Belgium; 9) Inflammatory Bowel and Immunobology Research Institute, Cedars Sinai Hospital, Los Angeles, CA; 10) The Meyerhoff Inflammatory Bowel Disease Center, Johns Hopkins Hospital, Baltimore, MD.

Colonic IBD is a significant risk factor for colorectal cancer (CRC) and high-grade dysplasia (HGD). Extensive involvement of the colon, the co-existence of primary sclerosing cholangitis (PSC), disease activity, family history of CRC and disease duration have been shown to increase the risk for CRC. However, no specific genetic association has been repeatedly shown to increase risk for CRC in IBD. We aimed to identify genetic associations with CRC/HGD in patients with colonic IBD using data from the ImmunoChip in a large multi-national cohort of patients with colonic IBD.

Membmbers of the International IBD Genetic Consortium (IIBDGC) were asked to identify patients with colonic IBD who developed CRC or HGD, verified by pathology. Demographic and clinical data were also collected. For each HGD/CRC case 1-2 controls were matched (by IBD subtype, disease duration, endoscopic extent (Montreal), age and sex). We aimed to identify genetic associations with CRC/HGD in patients with colonic IBD using data from the ImmunoChip in a large multi-national cohort of patient with colonic IBD.

Overall, 585 colonic IBD cases (390 UC, 171 CD, 19 IBDU) were reported, 774 had either HGD or CRC and 311 were matched controls. There were 58% male, 45% female, 15% current smokers. The mean age at diagnosis was 33. Disease duration to CRC/HGD diagnosis was significantly longer than length of FU in controls (230 months vs. 180 months, p=0.005). There was no significant difference in proportions of patients with extensive disease between the groups. However, HGD was more common in cases vs. controls (11.3% vs. 2.9%, p=0.004). Genetic data from IChip were available on 120 cases, 227 controls (59.3% of the total cohort; 50% males). Only Caucasians were selected for analysis. There were no significant genetic associations.

In this initial analysis, no genetic associations could be found related to risk of CRC/HGD in IBD. The IIBDGC will continue to gather cases to increase the sample size.

784W
Assessing genetic association and gene-gene interaction between PTPN22 and CSK for SLE susceptibility in Asians. X.R. Kin-Howard, C. Sun, A. Adler, S.C. Bae, S.K. Swapam. 1) Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea.

Regulation of effector T cell responses is essential for maintaining immune homeostasis and tolerance. Protein tyrosine phosphatase non-receptor type 22 (PTPN22) negatively regulates T cell activation associated with multiple autoimmune diseases, including systemic lupus erythematosus (SLE). It encodes intracellular lymphoid-specific phosphatase (Lyp), which has a prominent role in lymphocyte activation and differentiation. While functional coding variant (R620W, rs2476601) is robustly replicated in European derived populations the risk variant is absent Asian populations. Regulatory variant (rs34933034) of CSK, encodes protein c-Src tyrosine kinase (Csk), is associated with SLE in European-Americans. Lyp physically interacts with Csk during activation of downstream Src kinases, which may increase susceptibility to SLE. Our goals are to (a) assess these variants in an Asian population (Korean) using imputation-based association, (b) explore gene-gene interactions, and (c) identify association with SLE clinical manifestations (i.e. lupus nephritis) and autoantibodies (i.e., anti-Ro, anti-La, anti-dsDNA). We genotyped SNPs from PTPN22 and CSK in 1710 SLE cases and 3164 controls from Korea. Next, a comprehensive imputation-based analysis was used to add out-of-study GWAS controls and increase SNP density using 286 Asians available from 1000Genomes Project. Case-only analysis used specific SLE clinical manifestations as the phenotype. Since we observed significant association for PTPN22 and CSK SNPs we tested for potential genetic interaction between SNPs. This data analysis is in progress. We analyzed 138 SNPs from PTPN22 and 57 SNPs from CSK. As expected, known PTPN22 SNP rs2476601 was monomorphic in our Korean cohort. The most significant PTPN22 SNPs were in complete LD (rs21=1), intronic rs12746551 and rs39965092 (P=9.2x10^-6, OR=2.09). In CSK published SNP (rs54933034) was not significant (P=0.30) in Koreans. The most significant SNP was intronic rs4886627 (P=4.0x10^-4, OR=1.15). We observed suggestive interaction (P=0.07) between intronic SNPs rs39965092 (PTPN22) and rs11635664 (CSK). Clinical phenotype analysis is ongoing. Previously reported functional variants rs2476601 (PTPN22) and rs34933034 (CSK) were monomorphic and non significant, respectively. However, other potential SNPs from both genes could be associated with SLE. Additionally, we observed suggestive SNP-SNP interaction.
Correlation between SNPs within the MHC region and immune responsiveness to childhood vaccinations. Y.A. Taizhanov\textsuperscript{1}, B. Yucesoy\textsuperscript{2}, V.J. Johnson\textsuperscript{3}, N.W. Wilson\textsuperscript{4}, R.E. Biagini\textsuperscript{1}, M.I. Luster\textsuperscript{5}, M.M. Barrameda\textsuperscript{1}. 1) Center for Computational Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Toxicology and Molecular Biology Branch, CDC/NIOSH, Morgantown, WV; 3) BRT-Burleson Research Technologies, Morrisville, NC; 4) Department of Pediatrics, School of Medicine, University of Nevada, Reno, NV; 5) Division of Respiratory Diseases Studies, CDC/NIOSH, Morgantown, WV; 6) Biomonitoring and Health Assessment Branch, CDC/NIOSH, Cincinnati, OH; 7) Toxicology and Environmental Health, Research Triangle Park, NC. 

The present study investigated association between genetic variability in major histocompatibility complex (MHC) and immune response to childhood vaccinations. 141 healthy infants who have been immunized, according to standard vaccination scheme, with hepatitis B (HBV), 7-valent pneumococcal conjugate (PCV7), and diphtheria, tetanus, acellular pertussis (DTP) vaccines were participated in the study. Using multiplex immunassays we have assessed vaccine specific antibody responses and total serum immunoglobulin levels (IgM, IgA, IgG and IgG subclasses). We collected genotype information using illumina Goldengate MHC panels. Our analysis showed association between single nucleotide polymorphisms (SNP) within MHC region and vaccine specific antibody responses, serum levels of immunoglobulins (IgG, IgA, IgM) and IgG isotypes (IgG1-IgG3). Functional annotation of SNPs using RegulomeDB showed that both significant and correlated SNPs regulate the expression of a group of genes involved in antigen processing and presentation, including HLA-A, HLA-A, HLA-DRB5, HLA-DOA1, HLA-DOB1 and TAP-2. The results suggest that immune response to common childhood immunization is affected by genetic variations within MHC region, which in turn may influence vaccine efficacy. These novel regions within MHC may serve as candidates for future genetic and mechanistic studies.

The CTLA4 gene encodes a T-cell receptor, namely, cytotoxic T-lymphocyte antigen 4 (CTLA4), which is involved in controlling the proliferation and apoptosis of T lymphocytes. Among them, rs3087243 G>A and rs231775 A>G are the most associated markers with autoimmune diseases. In this study, we investigated whether the CTLA4 gene was associated with severe bacterial infections in Han Chinese children. Subjects and Methods Patients The patients were 142 unrelated children (54 boys, 88 girls) with severe bacterial infections (SBI) (bacterial septicemia, meningitis, pneumonia, or cellulitis). Their age at diagnosis was 4.9 ± 3.6 years (range 0.0-17.7 years). Among them, 55 were infected by pneumococcus. The control subjects were 92 healthy infants. They included hospital personnel and individuals who undergone routine health examinations or minor surgery. All patients and control subjects were ethnic Han Chinese in Taiwan. Our institutional review board approved this study, and all subjects or their guardians gave informed consent. CTLA4 genotyping We genotyped -318 C/T, +49 A/G, and CT60 polymorphisms of the CTLA 4 gene by PCR-restriction fragment length polymorphism (PCR-RFLP) using Msel, BstEII, and NcoI enzymes, respectively as we have reported. The -318 C/T polymorphism was amplified with primers 5'-CAC CAC GT-3' and 5'-CTG CTG AAA CAA ATG AAA CCC-3' resulting in a 153 bp product. The CT60 polymorphism was amplified with primers 5'-AAG GCT CAG CTG AAC CTG CTC-3' and 5'-GGG CGA CCT CCA CTG AAC-3' resulting in a 216 bp product. Statistical analysis Genotype, allele and carrier frequencies of the -318 C/T, +49 A/G, and CT60 polymorphisms of the CTLA4 gene were determined by direct counting. Agreement with Hardy-Weinberg equilibrium (HWE) was tested for genotype frequencies of the controls using haplovlew 4.2 Results The -318, +49, and CT60 polymorphisms of the CTLA4 gene were in Hardy-Weinberg equilibrium in controls and in patients. Statistical analysis revealed that no significant difference in the frequencies of genotypes, alleles, carriers, or haplotypes of the 3 SNPs of the CTLA4 gene between patients and controls. The results were similar when patients with pneumococcal infections were analysed separately. Discussion More patients must be collected.
Multiple eQTLs of TNFSF8 are associated with pathological immune responses in leprosy. V.M. Fava1,2, A. Cobat1,2, V.T. Nguyen1, N.N. Ba2, M. Orlova1, M. Vezinet1, M. Pintado1, A.C.P. Lalani3, M.M.A. Stevens1, A. Alcais1,2, E. Schurr1,2, K. Thangaraj1, A.Mishra1,2, 3 1) The McGill International TB Centre, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 2) Departments of Human Genetics and Medicine, McGill University, Montreal, Quebec, Canada; 3) Hospital for Dermato-Venerology, Ho Chi Minh City, Vietnam; 4) Lauro de Souza Lima Institute, Bauru, São Paulo, Brazil; 5) Tropical Pathology and Public Health Institute, Federal University of Goiás, Goiânia, Brazil; 6) Core for Advanced Molecular Investigation of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 11065, USA.

A genome-wide association study (GWAS) in Chinese leprosy patients identified six genes associated with leprosy. While five of these genes have been validated, one gene, TNFSF15, has resisted all attempts of replication by independent groups. While evaluating the clinical characteristics of the leprosy GWAS sample, we noted that more than 85% of the patients presented with disabilities. Leprosy type 1 reversal reactions (T1R) are the major cause of nerve damage in leprosy patients due to excessive immune responses directed against Schwann cells. We hypothesized that the lack of validation of TNFSF15 was in truth due to an association with the endophenotype T1R. To test our hypothesis, we studied SNPs in the TNFSF15 genomic region as risk factors for T1R in three independent populations: a family based sample from Vietnam and two case-control samples from Brazil, comprising a total of 1768 subjects. For the family-based design, we used a set of leprosy families without T1R as controls. In Brazil, the cases were T1R patients while the controls were leprosy patients without T1R. In total, we tested 19 SNPs across the Vietnamese sample, eight SNPs were strongly associated with T1R (p < 0.001). Unexpectedly, the majority of SNPs were in four bins that extended toward the neighboring TNFSF8 gene and not to the physically closer TNFSF5 gene. Indeed, of the 19 cis-eQTLs for TNFSF15 described in 14 GWAS databases, 17 had p < 0.05 in support of association with T1R. Multivariate analyses revealed independent association of two bins that carried numerous eQTLs for TNFSF8. The bins associated with T1R in the Vietnamese sample were found in the Brazilian samples. Yet, these bins were validated in Goiânia, four SNPs of an eQTL bin displayed p < 0.05 while in Rondonópolis eleven distinct SNPs presented p < 0.05 in favour of association with T1R. Taken together, our results identify a set of genetic risk factors for T1R in Vietnam and Brazil. These risk SNPs are eQTLs for TNFSF8 that despite of being in the same SNP bin may independently act on TNFSF8 expression levels. Our results illustrate the importance of taking in account endophenotypes in genetic association studies.
and identify the causative variant(s).

variants with predicted P values $1 \times 10^{-7}$. We next plan to perform de-

of 2,008 cases and 2,018 controls. We found we could accurately recapitu-

data had >99% concordance, emphasizing the confidence of the sequencing

SNPs that overlapped with the previous Illumina Human 660W BeadChip

0.05. The smallest P value was found 23 variants exceeding a nominal P

70% were novel. We performed association analysis on this dataset and

targeted re-sequencing of this region of interest (ROI) in 240 DSS cases

high linkage disequilibrium and hence, could be merely tagging causal var-

syndrome (DSS), we identified common SNPs within the PLCE1 gene to

gapore.

In our recent genome wide association study (GWAS) of dengue shock

system (DSS), we identified common SNPs within the PLCE1 gene to

Identification of genetic susceptibility loci for dengue shock syndrome

at the PLCE1 gene using a targeted re-sequencing approach. E. Png\textsuperscript{1},

T. N. Chau\textsuperscript{2}, N. M. Nguyen\textsuperscript{2}, D. T. Kien\textsuperscript{3}, N. T. Quyen\textsuperscript{2},

D. T. Trung\textsuperscript{2}, J. Pang\textsuperscript{2, 3}, B. Wills\textsuperscript{2, 4}, N. Van Vinh Chau\textsuperscript{4},

C. C. Khor\textsuperscript{1, 3, 6}, C. P. Simmons\textsuperscript{2, 4}, M. L. Hibberd\textsuperscript{1, 3}. 1) Infectious Diseases, Genome Institute of Singapore, Singapore; 2) Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam; 3) School of Public Health, National University of Singapore, Singapore; 4) Centre for Tropical Medicine, University of Oxford, Oxford, United Kingdom; 5) Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam; 6) Department of Paediatrics and Department of Ophthalmology, School of Medicine, National University of Singapore, Singa-

gapore.

The genetic polymorphisms associated with functional variations in Toll-like

receptors (TLRs) are candidates for such risk factors of infection/infectious
diseases. We conducted a population-based study of innate immune pheno-
type analysis for the better understanding of genetic predisposition to infec-
tious diseases. 1,999 newborn babies, who were born in Khanh Hoa Provin-
cial Hospital, a principal medical facility in Nha Trang City, central Vietnam
from May 2009 to May 2010, were enrolled to a longitudinal study. Genomic DNA samples were prepared from umbilical cord blood collected at the
entry. The events of severe diseases were passively but comprehensively
identified by admissions to the pediatric ward of the same hospital in which
they were born. The cohort children were invited to medical check at commu-
nity health stations at 24 month of age and venous blood samples were
collected upon their guardians’ informed consent. A simple procedure of
small-scale whole blood culture was established; fresh whole blood was
mixed with cell culture medium at 1:9 ratio and incubated at 37 degree in
the presence of LPS, Pam3CSK4 or other chemical stimulants of immune-
competent cells. Twenty hours later, the cells were collected for RNA prepa-
ration. The levels of induced mRNA expression of cytokine genes such as
IL8 and those of internal standards such as RPLP0 were quantified by real
time RT-PCR assay using Fluidigm Biomark platform. Among 1,999 cohort
children, 1,494 were responded to the invitation to 24-month medical check
and 1,348 blood samples were successfully collected. The first 368 blood
samples were examined for gene expression phenotype by the small-scale
whole blood culture. The incidence of severe acute lower respiratory tract
infection (ALRI) was 77.3/person/year (231 events in 1,494 persons in two
years). A major haplotype of eight SNPs in the TLR10-TLR1-TLR6 region
(4p14) was associated to ALRI as well as basal and inducible expression
of several immune-related genes including chemokines CCL3 and IL8, and
TLR1 itself. We observed a significant nonadditive interaction between
genetic polymorphism (a dose-effect of SNP haplotype) and gene expression
phenotype (classification of multiple chemokine mRNA levels by cluster
analysis) to the establishment of the risk for ALRI (hazard ratio of approx. 3).
795W

Systematic characterization of allelic architecture at the non-coding type 2 diabetes locus 9p21 using haplotype analysis in complete sequence data, imputation, and functional testing. V. Agarwala1,2,3, J. Flannick1,2, A. Morris1, M. Rasmussen2, M. Majahan3, H. Kang1, C. Fuchsberger1, M. Kellis1, M. Boehnke1, M. McCarthy4, D. Altshuler5,6, The Go-T2D Consortium. 1) Program in Medical & Population Genetics, Broad Institute, Cambridge, MA; 2) Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA, USA; 3) Program in Biophysics, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA, USA; 4) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6) Computer Science and Electrical Engineering Department, MIT, Cambridge, MA USA; 7) Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; 8) Department of Genetics, Harvard Medical School, Boston, MA, USA.

The causal variant(s) underlying signals discovered in genome-wide association studies for complex traits are, in most cases, unknown. The allelic architecture at these loci - the number, frequencies, and effect sizes of causal mutations - is not yet understood, in part because GWAS directly tested only common SNPs. The non-coding chr9p21 GWAS locus (near CDKN2A/B) harbors one of the strongest known signals for type 2 diabetes (T2D). Interestingly, multiple independent signals at this locus create a three-tiered (risk, neutral, and protective) haplotypic association. The availability of whole-genome sequencing data enabled us to test, for the first time, three current hypotheses about allelic architecture at 9p21: (1) that common variant(s) causally modulate risk of T2D, (2) that rare variants create ‘synthetic’ common variant associations, and (3) that rare variants (individually or in aggregate) have effects on T2D, independent of the common signals. To systematically test these hypotheses, we identified all variants, common and rare, across a 9Mb region around the 9p21 locus in ~2800 whole-genome sequenced Europeans. We confirmed the known haplotypic association and enumerated all variants present on the risk/protective haplotypes by integrating estimated phase data with information from ancestral recombination graphs. We tested variants, individually and in combination, (a) association to T2D, and (b) ability to explain the common signals. We performed imputation in the sequenced sample, and also imputation into ~35K samples densely genotyped on Metabochip. Observed genetic data place an upper bound on the contribution of rare variants at 9p21. Moreover, we reject the synthetic association hypothesis: we find no set of rare variants (excluding the rare variant that drives the signal in the haplotypic association, suggesting that causal variants underlying this GWAS signal are indeed common. After imputation, the credible set of variants with >99% probability of explaining the signal included only 1 common variant on the risk and 4 common variants on the protective haplotypes. Lastly, we conducted functional screens for regulatory activity across the 9kb region in multiple cell types. These experiments identified a sub-fragment which has beta cell-specific activity, and which includes the sites of the top candidate causal variants. The approach described here represents a general method for elucidating the causal architecture at disease-associated loci.

796T

A reliable algorithm for identifying modifier genes and the mutants that carry them in family-based studies of complex disorders. P. Belleau1, S. Dubois1, S. Desjardins2, R. Arsenault1, E. Shink2, J.L. Ancill2, G. Côté3, M.A. Walter1, M.A. Amyot4, V. Raymond5. 1) Neurosensory Systems Laboratory, CHUL Research Center (CHU de Québec), Quebec City, QC, Canada; 2) Ophthalmology, Université Laval, Quebec City, QC, Canada; 3) Medical Genetics, University of Alberta, Edmonton, AB, Canada; 4) Ophthalmology, Université de Montréal, Montreal, QC, Canada.

Our goal is to characterize gene-gene interactions implicated in open-angle glaucoma, a genetically complex disorder of the eye. Using a pedigree-based genome-wide linkage strategy and 154 heterozygotes for the autosomal dominant glaucoma myocilin R423E mutation, we recently mapped at 20p13, a modifier locus for variability of ages-at-onset (AAO). To characterize this modifier, we designed a pedigree-based algorithm that optimizes the identification of reliable double-mutants for genome-targeted sequencing. Our three-stage algorithm exploits a database obtained after genotyping 184 pedigree members with microsatellite and SNP markers covering the locus. The first stage is the identification of alleles associated with the modifier. The 2nd stage is the identification of double-mutants which simultaneously carry MYOCR423E and 1 marker allele associated with extreme values for AAO within the distribution of all AAO. The 3rd stage is haplotypes building starting with the alleles defined in 2. More formally, we first define as the kinship coefficient (Φ(X,Y) ≥ 0.0625 meaning that X and Y are people closer or equal to first degree cousins). Second, let D be a double-mutant carrying MYOCR423E and the allele A under study. The neighborhood of D is then defined as X, the set of MYOC carriers not carrying A where Φ(D,X) ≥ 0.0625. We compute for each MYOCR423E double-mutant the difference between this double-mutant AAO and the median of the AAO of his/her neighborhood. For all double-mutants, we then evaluate the median of the preceding difference and call this median I. We next compare I with the value Ip (p: permutation) calculated from permutations of AAO within the respective neighborhood of each double-mutant of the pedigree. The alleles for which the proportion of Ip > I is max. or min. are considered associated with the modifier. From the pool of MYOCR423E heterozygotes which carry the associated alleles (As), we only keep for sequencing the double-mutants who are at the extreme of the distribution of the AAO. Conversely, we keep as controls, MYOCR423E carriers who do not harbor the As alleles at the opposite extremes. Finally, we test for associated haplotypes starting at the As alleles. In conclusion, we designed a reliable pedigree-based algorithm that optimizes the identification of equivocal double-mutants that participate in gene-gene interactions. This algorithm can be applied to other quantitative traits and to other types of allelic markers for complex traits.
Nearly 70 type 2 diabetes (T2D) susceptibility loci have been identified to date, but the underlying causal variant(s) are not well characterized. One approach to fine-map these loci is to combine genome-wide association (GWA) data from diverse populations to increase sample size and leverage differences in linkage disequilibrium (LD) between ethnicities. We previously reported trans-ethnic fine-mapping analyses undertaken in five T2D loci: CDKAL1, CDKN2A/B, FTO, IGF2BP2 and KCNQ1, which demonstrate strong signals of association and differential patterns of LD across ancestry groups. In the current analysis, we increased the sample size to 64,727 (21,997 T2D cases and 42,730 controls), representing five ancestry groups (European, East Asian, South Asian, Hispanic and African American), and combined GWA data imputed up to the 1000 Genomes March 2012 reference panel of 10,92 individuals from diverse ancestry groups. Using trans-ethnic meta-analysis, we defined ‘credible sets’ of SNPs at each locus that have 99% probability of containing the causal variant, and evaluated the evidence for multiple independent association signals via conditional analyses. None of the lead SNPs at any of the loci showed substantial evidence for heterogeneity in allelic effects between ethnicities, suggesting that the underlying causal variants are shared across ancestry groups and are amenable to trans-ethnic fine-mapping. By comparing credible sets of SNPs identified through ancestry specific and trans-ethnic meta-analyses, we saw improved resolution in fine-mapping at all five loci by combining data from diverse populations. The resolution was most improved at CDKAL1, where the 99% credible set included 5 SNPs mapping to 12.9kb, compared with 15 SNPs identified in the five loci in the trans-ethnic fine mapping. When we localized to a region < 10Kb, there were two distinct signals defined by rs10961248 (Pcond=7.2x10^-10), rs234864 (Pcond=8.4x10^-8), and rs231353 (Pcond=1.7x10^-11). At CDKN2A/B, there are two distinct signals defined by rs231353 (Pcond=1.7x10^-11). We apply this approach to 66,849 QC’ed European samples from the International IBD Genetics Consortium’s immunochip project (18,603 CD, 14,307 UC and 33,938 controls). Out of 186 high-density regions in the immunochip, 76,862 variants in 110 regions have overlaps with the reported risk loci. These variants were imputed to the 1000 genome reference. 237,736 variants passed the imputation quality threshold and were used in the fine mapping.

We found that known causal variants were usually mapped as the only or the top variant in the credible set. For example, 3 NOD2 causal variants, fs1007insC, R702W and G908R were the only credible variants for the first 3 independent associations respectively. Out of the 110 fine-mapped regions, 50 have multiple independent associations and 78 have genome-wide significant (GWS) primary associations. Here we report fine-mapping results for these 78 associations. We reduced 11 associations to a single independent association. There are 13 associations that were localized to a region < 10Kb (28 of them have ≤ 5 variants). 13 associations have coding variants in the credible sets and 2 of them are the only variant. 9 out of the 11 single-variant credible sets are non-coding, suggesting high confident non-coding causal variants. We observed that in general, sizes of the credible set decrease with sample size, or the significance of the association signal.
800F

Autism spectrum disorders (ASDs) are a group of common and very highly hereditable neurodevelopmental impairments, characterized by the presence of three main behaviours: alteration of social interactions, abnormal language development, and the presence of stereotypic behaviours. These disorders are highly heterogeneous and in some cases its cause can be explained by highly penetrant de novo mutations; or by the impact of various mutations in a particular individual, each with a moderate effect, that act in synergy producing the autistic behaviours. It has been shown that a variety of genes are associated with ASDs and interact with the neuroligins and neurexins in the glutamatergic synapsis, being Shank3 one of these genes. The relationship between Shank3 and ASDs has been demonstrated with diverse genetic and functional studies that suggest a role of this gene in the pathogenesis of autism and in the formation of the dendritic spines. To recognize the importance of variations on this gene on our population, we made a screening of variations in a group of ten Colombian individuals with a diagnostic of ASDs and we determined the possible pathogenic effect of each of these changes. We found 9 variations, of which just one was found in the exonic region. As far as the intronic variations two changes occurred (rs113644328, MAF=0.005 and rs148315568, MAF=0.004) and stood out because of their low frequency, regardless of the bioinformatic analysis of these two changes did not show any significant importance. A missense mutation in exon 19 (rs16729471; p;Ala721Thr) with low frequency (MAF=0.012) was found in heterozygosity in one individual. Based in the structure and conservation of this position, the bioinformatic analysis suggest that this change produces a possible reduction in the stability of the protein; and the evaluation of different scenarios indicate that the effect might possibly be due to changes in the protein phosphorylation. Given the heritability of this change from an apparently healthy mother and the presence of the variation in a control population of other study, it can be presumed that even though this change is not responsible of ASDs in this individual, it might have a small effect which in synergy with other changes produces the autistic phenotype.

801W

Substantial progress has been made in identifying susceptibility variants for age-related macular degeneration (AMD), one of the most common causes of adult onset blindness. Major autosomal risk factors include Complement Factor H (CFH) Y402H (rs1061170) and Age-related Maculopathy Susceptibility-2 (ARMS2) rs10490924. Non-autosomal variants such as those in the mitochondria also contribute risk toward AMD. To further characterize the association between AMD risk and mitochondrial variants, we as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed the National Health and Nutrition Examination Surveys (NHANES) III and 2007-2008 for targeted genotyping of 60 mitochondrial SNPs. Fundus photography was available for participants >40 years. AMD cases were defined among older adults >60 years that presented with early/late AMD. Controls were defined among older adults >60 years without evidence of AMD. A total of 332, 37, 47 cases and 1500, 430, 270 controls for non-Hispanic whites, non-Hispanics blacks, and Mexican Americans, respectively, were available for study. We performed tests of association between AMD and mitochondrial variants using logistic regression assuming a dominant genetic model unadjusted and adjusted for age, sex, body mass index, and smoking status (current vs. none). Tests were performed for individual SNPs and by haplogroups J, T, and U. In non-Hispanic whites at a liberal significance threshold p<0.05, two variants were associated with AMD: rs28357682 (MT-CYTB) (OR=2.17, 95% CI=1.05-4.52; padj=0.04) and mt16362 (OR=1.81, 95% CI=1.00-3.00; padj=0.02). In our Mexican Americans, mt16111 was associated with AMD (OR=2.90, 95% CI=1.17-7.18; padj=0.02) and mt16362 (OR=2.48, 95% CI=1.01-6.15; padj=0.05). After adjustment for covariates, no mitochondrial variant or haplogroup was associated in non-Hispanic blacks, and no association generalized across all three race/ethnicities. The lack of associations in non-Hispanic blacks reflects the relatively low sample sizes and subsequent lower power to detect associations for this disease in African-descent populations, highlighting the need for further genetic association studies in diverse populations.
Bayesian conditional analysis of 180,000 individuals for migraine reveals local genetic substructure and nine additional genes. V. Anttila\textsuperscript{1,2,3,4}, H. Huang\textsuperscript{1,2}, N. Eriksson\textsuperscript{1}, E.M. Neale\textsuperscript{1,2}, M.J. Daly\textsuperscript{1,2}, A. Palotie\textsuperscript{2,3,5,6,7}, 1) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 4) 23andMe, Mountain View, CA, USA; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK.

Migraine is a common, episodic neurovascular disorder with significant heritability. Recently, our Consortium conducted a GWAS meta-analysis of 23,000 migraineurs and 95,000 controls, uncovering genome-wide significant association at 12 loci (Anttila et al, Nature Genetics, in press). In addition, we identified 134 additional loci with genome-wide suggestive p-values (less than 1 x 10^-5). We set out to apply a new Bayesian analysis framework to the dataset, allowing for a rapid testing of SNPs conditional on local genotypes from summary-level data. Additionally, we wanted to identify new genetic loci for migraine, using existing data and new samples from 23andme, for a total of 34,000 migraineurs and 147,000 controls. SNP data from the 146 implicated loci were analyzed in a meta-analysis of the previous cohorts and the new samples. In the conditional analysis, six new significant genes (TMEM51, MBOAT4, SUV39H2, PRKG1, HPSE2, TMEF91) were identified. All genes are from loci with p-values <1 x 10^-6 in the single-marker analyses. In addition, multiple signals were identified at four previously reported loci (PRDM16, MEF2D, TRPM8, LRP1), where accounting for the multiple signals increases the association p-values by several orders of magnitude (e.g. LRP1 p-value decreases from 3.94 x 10^-19 to 1.41 x 10^-26). In the expanded meta-analysis including the additional samples, four new loci showed significant association to migraine (FGF6, ASTN2 [distinct locus from previously reported], PRKG1 [significant also in the conditional analysis], MRVI1). At the MRVI1 locus, significant association is observed to a missense variant for the first time in migraine. In addition, MRVI1 interacts directly with PRKG1 in regulating many cellular functions; PRKG1 is a master regulator of many pathways involving in nociception and is a well-known regulator of intracellular calcium in neuronal cells, an emerging mechanism in neuropsychiatric diseases. We report the analysis of common variants influencing the genetic susceptibility to migraine, including nine new loci, as well as multiple local independent associations underlying previously reported loci. The newly identified loci include the first common missense variant in common forms of migraine. The implication of the genes at these loci, especially the PRKG1/MRVI1 pair and their biological pathways and mechanisms, make these findings a potentially interesting addition to our knowledge of the genetics of migraine.
A polymorphism of Interleukin-13 (IL13) is associated with susceptibility to food allergy in the Japanese population. T. Hirota1, M. Tamar1, M. Kubo2, S. Satoh3, M. Ebisawa4, T. Imas4, M. Sakashita4, S. Fujieda4,1) Laboratory for Respiratory and Allergic Diseases, IMS, RIKEN Yokohama Laboratory for Respiratory and Allergic Diseases, IMS, RIKEN, Yokohama, Japan, 2) Laboratory for Genotyping Development, IMS, RIKEN, Yokohama, Japan, 3) Clinical Research Center for Allergy and Rheumatology, Sagamihara National Hospital, National Hospital Organization, Sagamihara, Japan; 4) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan; 5) Department of Otorhinolaryngology-Head & Neck Surgery, Faculty of Medicine, University of Fukui, Fukui, Japan.

Background: The prevalence of food allergy (FA) has increased over the past two decades in Japan. FA is an adverse health effect arising from a specific immune response that occurs reproducibly on exposure to a given food. Recent GWASs have shown that IL13 locus on chromosome 5q31 is associated with bronchial asthma and atopic dermatitis. But influences of genetic variations in the IL13 gene on susceptibility to FA in the Japanese population are unclear. Objective: To investigate the association between polymorphisms of IL13 gene and FA and FA related phenotypes in the Japanese population. Methods: We resequenced the IL13 region by PCR-directed sequencing. Four tag SNPs were selected using the Tagger algorithm and genotyped by Taqman and Invader methods. We performed an association study of FA using a total of 603 subjects with FA and 938 controls in the Japanese population. Among the FA cases, a total of 182 subjects have bronchial asthma. Results: We identified a total of 17 polymorphisms including a non-synonymous substitution (Arg144Gln). We found that rs12595686 was associated with FA under an allelic model (P = 0.0000081; OR, 1.37; 95% CI, 1.18-1.60). In further analyses of patient subgroups, we observed a strong association between rs12595686 and FA with bronchial asthma (P = 0.0000081; OR, 1.68; 95% CI, 1.34-2.12). Conclusion: Further analysis of the linkage and functional analyses are needed, our findings could help elucidate common genetic factors for bronchial asthma, atopic dermatitis, and FA.

No association between the neuropeptide Y gene polymorphisms and smoking habit in Japanese. M. Isomura, C. Matsuda, T. Nabika. Functional Pathology, Shimane University, Shimane, Shimane, Japan.

Purpose: Smoking induces various adverse effects such as chronic obstructive pulmonary diseases, lung cancer, as well as heart diseases. Quit smoking has been a large issue in public health to reduce smoking related diseases. Because nicotine in a cigarette produces activities in brain reward system, smoking habit has a strong potential for addiction. Recently, it has been shown that a NPY (neuropeptide Y) participates in brain reward system and genetic polymorphisms in NPY gene associates with smoking habit. To explore relationship between NPY and smoking habits, we have conducted genetic association study. Methods: 3,020 individuals attending health check examinations carried out at rural area in Shimane prefecture in Japan. Smoking history and status were self-reported. Mental status of participants was assessed from response to questionnaire. Three SNPs in NPY gene were genotyped by TaqMan method. Concentrations of NPY in serum were measured by using Neuropeptide Y EIA Kit (Phoenix pharmaceuticals). Results: Genotypes of three SNPs in NPY gene were successfully determined. Association between genotypes and smoking habits were examined by comparing allele frequencies of these SNPs according to smoking status (never, or ever smoker). However, no statistical differences were observed in allele or genotype frequencies between these groups (p=0.095). Although several reports showed that NPY concentrations in serum vary among genotypes of a SNP, rs16147, no difference was observed between CC genotype (1.52ng/ml) and TT genotype (1.57ng/ml) in our study. Discussion: In Japanese, polymorphisms in NPY gene showed no association with smoking habit, even though several confounding factors were adjusted. Discrepancy of published result and our result may be due to difference in ethnicity. Our results indicated that polymorphisms in NPY gene had little contribution to smoking habits in Japanese.
809F Common Variants Associated with Normal Tension Glaucoma Are Also Associated with Glaucoma in Exfoliation Syndrome. D. Wang1, B. Fan1, H. Levkovitch-Verbin2, I. Pasqualle1, J. Haines2, J. Wiggs1. 1) Ophthalmology, Massachusetts Eye & Ear Infirmary, Boston, MA; 2) Goldschleger Eye Institute, Sheba Medical Center, Tel-Hashomer, Israel; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Genome-wide association studies have identified several genomic regions associated with primary open angle glaucoma (POAG), including TMCO1 at 1q24, CAV1/CAV2 at 7q31, CAVH6D at 1q24, and TAC1/TC48 at 4q38. The lead SNPs rs284489 at 8q22 and rs1412829 at CDK2NB2 were significantly associated with EG in the US dataset (P = 0.02 and 0.01 respectively) and in the meta-analysis of both control datasets (P = 0.008 and 0.02 respectively). No evidence of association was found between the lead SNPs from the other POAG associated regions and EG in either population. There was no evidence of association between these lead SNPs and ES without glaucoma. These results suggest that genetic variants in the 8q22 and 9p21 regions may contribute to optic nerve disease in exfoliation glaucoma, and this may be risk factors for optic nerve disease related to glaucoma generally.

810W Alternately Spliced Isoform Expression of PTPN2 in Rheumatoid Arthritis. M. Houtman, K. Stokhelm, L. Paduchkov. Rheumatology Research Unit, Department of Medicine Solna, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden.

Background: Rheumatoid arthritis (RA) is a prototype autoimmune disease with unknown etiology. Despite extensive genetic association studies, the mechanism of disease in relation to putative function of found loci remains unclear. Expression analysis is one of the tools to link findings from genetic epidemiology studies with biological mechanisms of the disease. Variations within the gene encoding the intracellular protein tyrosine phosphatase non-receptor type 2 (PTPN2), which has been implicated in T cell receptor signaling, have recently been associated with a number of chronic inflammatory diseases, including RA. The PTPN2 gene codes for at least two alternatively spliced isoforms - TC45 and TC48, which differ in their C-termini and localize differentially to the nucleus and cytoplasm. In this study, we investigate mRNA expression of PTPN2 isoforms in peripheral blood cells of RA patients and healthy controls. Methods: We studied the relative mRNA expression of all PTPN2 splice forms combined and the specific PTPN2 splice forms expressed in whole blood samples from 75 RA patients and 77 healthy controls of Caucasian ancestry. The Mann-Whitney U-test was used to evaluate differences and P < 0.05 was considered to be significant. Results: There was no significant difference in total PTPN2 expression in whole blood samples from RA patients and healthy controls. However, the PTPN2 splice forms were differentially expressed in these samples. We saw an increase in expression for TC45 (median relative quantification range (IQR) 0.664); P = 0.0259) and TC48 (median ROI = 1.047 (IQR 0.529); non-significant) in RA patients. This difference was more pronounced when comparing the ratio of splice forms (P < 0.0001). Conclusions: Our study shows that PTPN2 expression of the canonical and the alternatively spliced isoforms of RA patients is different compared to healthy controls, although expression of total PTPN2 did not show this clearly. It is an example of expression analysis where detection of overall gene expression without considering transcription/translation diversity may generate false results. Our results suggest that the structural differences between the two isoforms might affect events in the pathogenesis of RA.

811T Comparative analysis of eQTL structures in chronic obstructive pulmonary disease (COPD) and Genotype-Tissue Expression (GTeX). T. Huang, Q. Long, B. Zhang, S. Yoo, J. Zhu, Z. Tu, the GTeX Consortium. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Expression quantitative trait loci (eQTLs) can help people to understand the mechanisms of disease associated SNPs as majority of them are in non-transcribed regions and presumably function through gene expression regulation. GTeX, like many existing eQTL data collections (e.g., seeQTL and Genevar), is based on tissues from relatively healthy individuals. It is unclear that to what extent eQTLs identified in these normal tissues could replicate in disease populations from which GWA studies were performed. Answers to this question are important and will provide guidance on the appropriate usage of GTeX eQTLs for GWAS SNPs interpretation and to shape the future direction for projects like GTeX. To help answer this question, we calculated eQTLs in all the nine tissues from GTeX with sufficient samples, and eQTLs from COPD lung tissues profiled from Lung Genomics Research Consortium (LGRC) project. Raw data from both data sets was processed in strict parallel to ensure fair comparison. The eQTLs from LGRC COPD lung, LGRC control lung, GTeX adipose, artery, heart, lung, muscle, nerve, skin, thyroid and whole blood tissues were compared from multiple angles. Several interesting results are found: first, eQTLs from GTeX lung share a great similarity with eQTLs from LGRC control lung compared to LGRC COPD lung. Second, among the nine GTeX tissues, lung is the one showing highest similarity with LGRC COPD/control lung with respect to eQTL structures. Third, 35.6% of eQTLs linked to COPD GWAS SNPs (or SNPs in high LD, r2 > 0.8) in LGRC COPD tissues are shared in GTeX. This indicates that genetic regulatory structures are partially conserved across disease and normal individuals. Interestingly, the GTeX tissue with most similar eQTL pattern for GWAS COPD associated SNPs with LGRC eQTLs is not lung but muscle. This may reflect certain pathological changes related the disease. In conclusion, we performed a comprehensive comparison of eQTL structures in disease and normal tissues and report our findings on the conservativeness of eQTL structures across disease/normal tissues, tissue specificities of these data with respect to their relevant disease and found some new insights to the complex disease systems and may help to better utilize GTeX for disease studies.

812F Phenotypic analysis of Peptidylarginine deiminase type 4 knock-out mice. A. Suzuki1, Y. Kochi1, H. Shodai2, K. Fujio2, R. Yamada1, 2, K. Yamamoto1, 2. 1) IMS, RIKEN, Yokohama, Japan; 2) Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Center for Genomic Medicine, Kyoto University, Kyoto, Japan.

Rheumatoid arthritis (RA) is well-known as an autoimmune disease and is a chronic inflammatory disorder characterized by the destruction of multiple joints. Many genome-wide association studies were performed and multiple RA-susceptibility loci and autoimmune-susceptibility loci have been identified. These studies suggested that multiple genes and its functions were related with disease causing and development. Previously, peptidylarginine deiminase type 4 (PADI4) was identified as a susceptibility gene for RA in a Japanese population by case-control association study (Ref 1). PADI4 is a member of the PADI gene family and converts arginine residue (peptidylarginine) to citrulline residue (peptidylcitrulline). PADI4 is highly expressed in bone marrow, macrophages, neutrophils and monocytes. Peptidylcitrulline is an interesting molecule in RA, because it is an antigen of ACPA and only PADs (translated protein from PADI genes) can provide peptidylcitrullines, via modification of protein substrates. To evaluate the importance of PADI4 gene in the progression of RA, we generated Padi4-/- DBA1J mice. We used Padi4-/- mice to show that PADI4 is significantly affected to progress of collagen induced arthritis (CIA), well known as an RA model animal. In Padi4-/- CIA mice sera, the concentrations of serum anti-CII IgM, IgG, and levels of inflammatory cytokines decreased significantly rather than in WT CIA mice. Furthermore, intracellular inflammatory cytokines in CD11b positive splenocytes from Padi4-/- CIA mice are also significantly lower than those from WT CIA mice. As the results, we suggested that PADI4 enhanced collagen-initiated inflammatory responses. 1) Suzuki, A. et al Nat Genet.34, 395-402 (2003).
813W
Defective PGRN gene expression in two patients with frontotemporal dementia (FTLD) with potential decreased mRNA stability due to exon 6 PGRN deletion. G.101349_101355delCTGCTGT was associated with frontotemporal FTLD. E. Vitale 12, A. Iuliano 1, A. Polverino 2, G. Milan 3, S. Papata 4, P. Sorrentino 5, A. Postiglione 6, V. Alesi 7, A. Novelli 8, G. Sorrentino 1, 2, 6, 8. 1) Cybernetics, CNR-National Research Council, Pozzuoli (NA), Napoli, Italy; 2) School of Movement Sciences (DISIST), University of Naples Parthenope, Napoli, Italy; 3) Geriatric Center "Frunolle" ASL Napoli 1, Naples, Italy; 4) CNR-National Research Council, Bioimaging and Biostructures, Naples, Italy; 5) Department of Clinical and Experimental Medicine, University of Naples Federico II, Napoli, Italy; 6) Neurogenetics and Treatment Hermitage Capodimonte, Napoli, Italy; 7) Mendel Laboratory, IRCCS Casa Sollievo della Sofferenza Hospital, Roma, Italy; 8) Medical Genetics Unit, San Pietro Hospital Fatebenefratelli, Rome, Italy.

Frontotemporal dementia (FTLD) is a neurodegenerative disease with hallmark deficits in social and emotional functions, characterized by genetic and clinic heterogeneity and progressive declines in behavior or language associated with frontal and temporal lobar degeneration. The phenotypic heterogeneity consistently reveals three different clinical manifestations such as a behavioral variant of FTD (bvFTD), primary progressive aphasia (PPA) and FTD overlapping with motor neuron disease (FTD-MND or FTD-ALS). It can also present with parkinsonian syndromes, progressive supranuclear palsy (PSP) and corticobasal syndrome (CBS). The progranulin gene (PGRN) was recently reported to cause tau-negative frontotemporal dementia linked to chromosome 17. Here we report a four-generation Southern Italian family segregating FTLD in four affected family members, two of them still alive. We collected these two cases (one male and one female) along with the unaffected UK brother. The two phenotypes are clinically heterogeneous for the behavioral variant (bvFTLD) in the males and for a Primary Progressive Aphasia (PPA) in the female. These individuals were analyzed by sequencing PGRN and other genes in search for a mutation. We found a g.101349_101355delCTGCTGT deletion in PGRN exon 6 CBS in the two affected individuals. The g.101349_101355delCTGCTGT deletion has already been described in two sporadic cases as causing a premature stop codon with a frameshift introduction and could be causing mRNA non-sense mediated decay resulting in PGRN protein haploinsufficiency. So, we have tained this mutation in about 50 healthy controls matched by age, sex and geographic regions using allele-specific PCR mutation-primers (ARMs) and found no deletions carried by these individuals. Moreover, a CGH assay performed on the DNA of the two patients showed no additional genomic aberrations. Quantitative RT-PCR of the PGRN gene using exon 7-8 forward and reverse primers was performed on mRNA obtained from WBC from the two patients and from controls. We found a defective PGRN gene expression in the two patients leading to a possible decreased mRNA stability. These results suggest that PGRN may be a possible modifier gene that could lead the FTLD phenotype. Further investigations will focus on understanding how PGRN mutations lead to neurodegeneration and, specifically, on development of FTLD in this family.

814T
Biomarkers and Perinatal Risk Factors in Autism Spectrum Disorder. H. Xu 1, 2, H. Wang 1, Q. Dai 1, A-Q. Zhou 1, M-R. Wu 1, X-Y. Wang 1. 1) Children's Developmental Medicine, Hubei Province Maternal and Children's Health Hospital, Wuhan, Hubei, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children's Health, China.

Autism spectrum disorder (ASD) has been given more and more attention with the increasing incidence rate worldwide. Despite the complicated behavioral observation scales that have been created for the diagnosis of ASD, few studies have focused on the use of biomarkers for early diagnosis of ASD, which may bring the identification and prevention of ASD to an earlier age. Here we have investigated the possible use of plasma secreted amyloid precursor protein alpha (sAPP) and brain-derived neurotrophic factor (BDNF) as the peripheral biomarkers in the diagnosis of ASD from more than 150 children with ASD. A sensitive ELISA method was used to detect plasma sAPP and BDNF levels. There were significantly increased levels of total sAPP in 67.5% sAPP-α in 57.4%, sAPP-β in 62.8%, BDNF in 51.22% of the clinically diagnosed autistic children. The association was statistically significant among the total sAPP (F = 59.4965, P = 0.0083), sAPP-α (F = 7.5036, P < 0.0001), sAPP-β (F = 1.7175, P = 0.0083) or BDNF (F = 10.0685, P = 0.0156) between case groups and control. Correlation analysis of sAPP and BDNF with age shows the total sAPP, sAPP-α, sAPP-β significantly positively associated with age (β were: -1.90849, -0.80611, -0.75111 respectively; P < 0.05 respectively). We have also used a self-designed questionnaire to screen risk factors for ASD and found 11 perinatal factors. We have therefore concluded that prenatal high-risk factors may have certain relationship with ASD; expressions of sAPP, sAPP-α, sAPP-β and BDNF in plasma are of different severity in children with different levels of ASD. It is a promising way for early diagnosis of autism by peripheral biomarkers both in plasma and cord blood, which will promote to build a new ASD diagnosis system.

815F
Analysis of DNA-protein interactions using nuclear extracts around the -224 A/G single nucleotide polymorphism in the neuropeptide receptor Y2 (NPY2R) gene in predisposition to hypertension. E. Albino, A. Nuñez, J. Dutil. Biochemistry, Ponce School of Medicine, Ponce, PR.

Previous work in animal models and human populations identified Neuropeptide Y receptor 2 (NPY2R) as a candidate gene for HTN. In two independent Japanese populations, the GG genotype single nucleotide polymorphism (SNP) located 224 bp upstream of the transcription start site was associated with an increased risk of HTN. In addition, the luciferase activity induced by the NPY2R promoter with the G allele in position -224 was reduced by 34% compared to the promoter containing an A in that position. The aim of this study was to assess the DNA-nuclear protein interactions surrounding the NPY2R-224 A/G SNP. Two versions of double stranded oligonucleotide probes (50bp) corresponding to the sequence flanking NPY2R-224 A/G were synthesized and incubated with different amounts of nuclear protein extracts for an electrophoretic shift assay (EMSA). The ENCODE project proposes CTFC as a functional element that binds to the -224 SNP in the NPY2R gene. We used CTFC antibody for a gel super shift assay. The DNA/nuclear protein reactions were loaded on a 4% non-denaturing polyacrylamide gel. Interestingly, the EMSA showed an allele-specific binding with the oligonucleotide containing an A nucleotide in -224 position. However, the oligonucleotide containing the G nucleotide in -224 position shows no DNA/nuclear protein interaction. Also, we were able to detect a super shift using the CTFC antibody. The data provides strong evidence for a functional role of NPY2R gene in predisposition to HTN.

816W

In the genomics era, one major challenge is the identification of causal functional SNPs and genes. Indeed, thousands of SNPs have been associated with various diseases through candidate gene and genome-wide association studies, but, only a few have experimental functional evidence to support their disease causality. This hinders the process of translating genetic association signals to targeted therapy. Our group has identified multiple genetic variants associated with several complex diseases, including HIV/AIDS, hepatitis B and C, and kidney diseases. Here, we employ multifaceted experimental approaches aided by bioinformatics tools to evaluate the functional effects of some of these individual genetic variants. Disease-associated SNPs functionality was assessed by real-time PCR to quantify gene expression in the Epstein-Barr virus-transformed lymphoblastoid cell lines (LCs) from patients, by disease target cell-based gene reporter assay to test the regulatory potential of SNPs, and by electrophoretic mobility shift assays (EMSA) to estimate the impact of SNPs on DNA transcription factor binding. Several SNPs in APOBEC3F, ZNRD1, CD4 and SERPINA1 of AIDS candidate genes were found to affect transcription factor binding, with patterns often differing in cell types. A ZNRD SNP in the U'-upstream region exhibited differential transcription factor binding pattern and altered promoter activity. Integration of these experimental data with in silico bioinformatics data provided a new insight into SNPs function relevant to the disease susceptibility. The biological function beyond genetic association may help us understand the operating mechanisms of genetic variants in causing various diseases. (Funded by the National Cancer Institute Contract HSN261200800001E).
817T
Generation of CLEC16A inducible knockout mouse as a novel model to study the pivotal role of NK cells in the pathogenesis of Type 1 Diabetes. M. Bakay1, M. Rankin1, S. Yoesum1, J. Kushmer1, H. Hakonarson2. 1) Center for Applied Genomics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 2) Texas Children’s Diabetes and Endocrinology Center, Texas Children’s Hospital, Houston, TX.

CLEC16A was recently found to be associated with several autoimmune disorders including Type 1 Diabetes (T1D). The C-type lectin domain family 16, member A (CLEC16A) gene encodes protein with C-type lectin domain structure, which makes it potentially related to the immune response. Although CLEC16A has no known function so far, the discovery of the C-type lectin gene as a diabetes gene, could eventually translate preventive therapy for the disease. We hypothesize that CLEC16A plays an important role in the immune system and that knockout (KO) mouse for CLEC16A will give valuable clues in this respect, particularly in exploring effects of CLEC16A knockdown that cannot be studied ex vivo. We generated CLEC16A mutant mouse by introducing mutation in exon 3 that leads to a frameshift and a premature STOP codon. Homozygotes for the mutation had a phenotype: small body size, weak tail, they may live to adulthood but die prematurely and do not breed. Heterozygotes have a normal phenotype and lifespan and are fertile. Homozygotes were approximately 50% the weight of their unaffected littermates at one month of age. They may live until adulthood, but die by 9 months of age, none of females has reproduced. Interbreeding heterozygotes produced smaller litters and less than the expected number of homozygotes. To overcome this problem and determine the role of CLEC16A in adult mice, we have generated inducible CLEC16A KO mice (Clec16aloxPloxPIns-CreERT). By our design, Osgene (Australia) has generated a conditional KO mouse for CLEC16A, by floxing exon 3. This introduces a frameshift inactivating all downstream exons. We have crossed Osgene mice with UBC-CreERT mice (inducible cre recombinase driven by the human ubiquitin C promoter). Tamoxifen treatment will inducibly activate Cre transgene which will lead to removal exon 3, flanked by loxP sites, in Clec16a gene. This event is predicted to result in a frameshift mutation, leading to the introduction of novel amino acids and early termination codon. Deletion will be confirmed in islets and other tissues by quantitative RT-PCR.

Since a whole body CLEC16A KO is a new mouse model, we are in a process of characterizing this model and studying if CLEC16A affects glucose metabolism, insulin secretion and tolerance. We also will determine the role of CLEC16A and its function in the progression of T1D by studying an inducible whole body knockout model with focus on NK cells.

818F
Genetic Polymorphism and Pathophysiology in Patients with Vitiligo. S. Chettiar1, K. Misty1, P. Agarwal1, A. Patel2, D. Jhala1, D. Umarigar1, R. Uppala1. 1) Department of Biotechnology, SRKl, surat, India; 2) Department of Dermatology, Civil Hospital, Surat; 3) Department of Zoology, Gujarat University, Ahmedabad; 4) Genepro Diagnose Research Laboratory, Navrangpura, Ahmedabad; 5) Department of surgery-Transplant, Gujarat University, Ahmedabad; 6) Department of surgery-Transplant, University of Nebraska Medical center, Omaha, NE, USA.

Vitiligo is an acquired, idiopathic, hypomelanotic, depigmentary disorder characterized by appearance of white patches resulting from the loss of functional melanocytes and melanin from the skin and mucous. It affects 1-4% of the world population, where as 0.5-2.5% in India with a high prevalence of 8.8% in Gujarat and Rajasthan states. Oxidative stress plays a vital role in etiology of depigmentation in skin by cellular loss, while catalase (CAT) is a proven enzymatic defense system, catalyzing break down of hydrogen peroxide. Altered activity of the enzyme and increased stress markers have been reported in vitiligo patients. To investigate CAT gene polymorphism association to vitiligo susceptibility, we investigated two CAT gene SNPs including promoter region rs7943316 (T/C) and exon 1 codification site of BstXI and Hinf I respectively, in 54 vitiligo patients and 45 healthy volunteers. Catalase activity from the serum of affected and normal individual showed a significant difference, thus supporting that oxidative stress might be involved in the pathophysiology of the disease. The genotype distribution and allele frequency promoter region are not significantly different between vitiligo patients and healthy controls. But, the exon 9 showed significant correlation between affected and healthy individuals. Although the haplotype of two polymorphisms also showedassociation with vitiligo. This study suggests possible association between the CAT gene and the vitiligo susceptibility.

819W
Allele-specific regulation of DISC1 expression by miR-135b-5p. I. Hovatta1,2, M. Rossi1,2, H. Kluipien1,2, M. Muona1,2, I. Surakka1,2, C. Ingel2, W. Hennah1,2, S. Ripatti1,2,3,4, M. Bakay1, M. Rankin1, S. Yoesum1, J. Kushmer1, H. Hakonarson2. 1) University of Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Wellcome Trust Sanger Institute, Hinxton, UK.

The Disrupted-in-schizophrenia-1 (DISC1) gene is a risk factor for neuropsychiatric phenotypes. MicroRNAs (miRNAs) are important regulators of protein coding genes. The miRNA-mRNA target recognition mechanism is vulnerable to disruption by DNA polymorphisms. We therefore investigated whether polymorphisms in the DISC1 3’UTR affect binding of miRNAs and lead to allele-specific regulation of DISC1. We identified four predicted polymorphic miRNA binding sites in the 3’UTR region of DISC1. We tested the effect of nine miRNAs on endogenous DISC1 expression in vitro by over-expressing these miRNAs in 293FT cells and measuring the DISC1 expression level by qPCR. To determine whether these miRNAs regulate DISC1 levels by targeting the predicted sites, we used a luciferase reporter gene assay. We cloned either the full length DISC1 3’ UTR or the miRNA binding sites into a luciferase expression vector and co-transfected them into 293FT cells with either the miR-135b-5p or the miR-559 precursor. We also investigated the putative allele-specificity of miR-135b-5p binding. The 293FT cell line is homozygous for the A allele of rs11122396, predicted to create a novel binding site for miR-135b-5p. Therefore, we created constructs with the A allele at rs11122396. Two of the nine miRNAs, miR-135b-5p and miR-559, significantly reduced endogenous DISC1 mRNA expression: miR-559 by 23.7% (p<0.009) and miR-135b-5p by 16.2% (p<0.039). In the luciferase assays expression from the DISC1 full length 3’ UTR construct was reduced 32.1% (p<0.003) by miR-135b-5p, and by 10.3% (p<0.03) compared to the construct floxing the 60 nt miRNA binding site and flanking sequences. In contrast, miR-599 over-expression did not affect the expression of DISC1. When investigating the putative allele-specificity of miR-135b-5p binding we observed that miR-135b-5p had no effect on the luciferase activity of either the full length 3’ UTR with G allele at rs11122396 (p=0.49), or on the construct with the –60 nt miRNA binding site insert (p=0.18), indicating that miR-135b-5p binding is specific to the derived (A) allele at rs11122396. Thus, the G allele may be functionally related to the DISC1 associated phenotype by abolishing regulation by miR-135b-5p, leading to elevated DISC1 levels.

820T
CELSR1 mutations are associated with human spina bifida. Y. Lei1, H. Zhu1, W. Yang2, M. Ross3, G. Shaw4, R. Finnell1,2. 1) Deli Pediatric Research Institute, The University of Texas at Austin. 1400 Barbara Jordan Blvd, Austin, TX, 78723; 2) Department of Chemistry and Biochemistry, College of Natural Sciences, The University of Texas at Austin, Austin, Texas; 3) Department of Pediatrics, Division of Neonatology, Stanford University School of Medicine, Stanford, CA; 4) Center for Neurogenetics, Brain and Mind Research Institute, Weill Cornell Medical College, New York, NY.

Spina bifida is one of the most common neural tube defects (NTDs), yet its complex etiology remains poorly understood. Recently, mutations in planar cell polarity (PCP) genes have been associated with NTDs, including spina bifida, in both animal models and human cohorts. We studied CELSR1, one of the PCP core genes in a California cohort to determine if CELSR1 mutations increase the risk of spina bifida. We sequenced the coding region of CELSR1 in 192 spina bifida infants. Novel and rare variants were genotyped in a control group of 190 ethnically-matched infants without malformations. Six missense mutations, absent in controls, were predicted to be deleterious by both SIFT and PolyPhen computational algorithms. Two TG dinucleotide repeat variations were detected in two spina bifida infants separately, but were not observed in controls. In functional analyses, the two TG dinucleotide repeat variants altered the subcellular localization of GTAG-containing CELSR1 protein in transfected cells. Both variants impaired interactions with VANGL2, and reduced the ability to recruit VANGL2 to the cell membrane. In total, 4.1% of spina bifida cases possessed deleterious or predicted-to-be-deleterious CELSR1 mutations. Our findings indicated that CELSR1 mutation is a risk factor for spina bifida.
821F
Functional impact of polymorphisms in the MMP3 and TIMP2 gene promoters in human nonsyndromic cleft lip/palate. A. Letra1,2,3, M. Zhao1, R.K. Silva1,2,3, A.R. Vieira4, J.T. Hecht1,2,3 1) Department of Endocrinology, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 2) Department of Pediatrics, Pediatric Research Center, University of Texas Health Science Center Medical School at Houston, Houston, TX; 3) Center for Craniofacial Research, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 4) Departments of Oral Biology and Pediatric Dentistry, University of Pittsburgh School of Dental Medicine, Pittsburgh, PA.
Matrix metalloproteinases (MMPs) and their inhibitors (TIMPs) are responsible for tissue remodeling during craniofacial development. Evidence from biological and human studies strongly support a role for MMP and TIMP genes as candidate genes for cleft lip/palate. We have previously shown the association of promoter polymorphisms in MMP3 and TIMP2 genes with cleft lip/palate. In this study, we assessed whether the previously associated promoter variants in MMP3 (-709 A/G) and in TIMP2 (-180 C/T) genes have functional implications and modulate gene transcription by changing the affinity of their gene promoters for transcription factors. Electrophoretic mobility shift assays, mass spectrometry analysis and luciferase reporter gene assays were performed for each gene variant. For the MMP3 -709 A/G variant, EMSA showed DNA-protein binding complexes specific for the A allele, in which 34/204 proteins identified by mass spectrometry analyses were specific for this allele. Luciferase assays showed that allele A has lower promoter activity than the G allele in the presence of a known functional polymorphism in the MMP3 promoter (-1171 S6A). For the TIMP2 -180 C/T variant, EMSA showed that both C and T alleles present potential binding sites for NF-Kappa B transcription factor, while super shift assays confirmed binding to NF-Kappa B with a higher binding affinity for allele C. Luciferase assays showed a 2.5 fold increase in promoter activity in the presence of allele T when compared to allele C. Taken together, these results show that variations in MMP3 and TIMP2 gene promoters may influence gene transcription and/or function with effects on craniofacial development. Our study provides new evidence implicating MMP3 and TIMP2 variants in the occurrence of cleft lip/palate.

822W
Indian Hedgehog (IHH) and LRBA: New genes for Hirschsprung disease. R.M.W Hofstra1, Y. Sribudiani1, R. Chauhan1, M. Alves1, C. Kockx2, T. van Essen3, R. Brouwer2, M. van den Hout2, W. van Ickelen2, 1) Department of Endocrinology, ErasmusMC, Rotterdam, Netherlands; 2) Center for Biomics, Erasmus Medical Center, Rotterdam, Netherlands; 3) Department of Biology, Emory University, Atlanta, USA.
Hirschsprung is characterized by the absence of enteric ganglia in a variable length of intestinal tract. A large, multi-generational Dutch family with five affected family members with HSCR, revealed linkage to 4q31.3-q32.3. As the family shows an autosomal dominant mode of inheritance with incomplete penetrance, we assume that the mutation in the linkage region is necessary but not sufficient to cause disease development. To identify the mutation in the linkage region, but also mutations elsewhere in the genome, we exome sequenced two patients. One variant rs140666848 (T > C) in exon 20 of the LRBA gene was found in the linkage region. In addition, missense mutations in RET (P398L) and IHH (Q50K) were identified in each patient, respectively. Functional analysis of the RET en IHH variants showed that both mutations give rise to a non functional protein. The LRBA variant is located downstream of MAB21L2, a gene which plays a role in the proliferation of enteric neural crest cells (ENCCs) during ENS development in Zebra fish. We hypothesize that this variant might regulate MAB21L2 expression in ENS. Luciferase assays showed that the region containing the variant does show enhancer activity. Currently we are testing whether the variant, when compared to wild type sequence, show significant differences on gene expression level, using luciferase assays. Our data shows that combinations of mutations, as expected, cause autosomal dominant disease with reduced penetrance. Furthermore, we show for the first time that IHH mutations can contribute to Hirschsprung disease.

823T
Genes located in type 1 diabetes risk loci are expressed in human islet and interact in functional networks. C.A. Bronsson, J. Stoerling, F. Pociot. Glostrup Research Institute, Glostrup University Hospital, Glostrup, Denmark.
Background: Genome-wide association studies of type 1 diabetes (T1D) have identified 50 susceptibility loci. Although most pin-pointed candidate genes have putative functions within the immune system studies in human pancreatic islets have shown expression for 60% of these genes. Assignment of candidate genes is often based on known biological function or location in relation to the most associated SNP in a locus which could introduce bias into functional studies as all genes in a locus potentially are equally good candidates. We investigated the expression and functional interactions of all genes located within T1D loci in human islets. Methods: A list of all 857 genes located within T1D GWAS loci were downloaded from t1dbase.org. Gene expression based on RNA-seq in human islets was determined from a published study by Eizirik et al. Only genes that had an expression level above 1 in at least five preparation in the control and/or cytokine-stimulated condition were considered expressed. Protein-protein interaction networks of expressed genes were constructed in DAPPLE. Results: Of 857 genes in T1D loci 336 were expressed in the islets representing 44 loci and 33 previously pin-pointed candidate genes. The expressed genes included 8 pseudogenes, three lincRNAs, two miRNAs and two snoRNAs. In 16 of the loci only a single gene was expressed, including nine candidate genes (STAT4, CENPW, TNFAIP3, SKAP2, GLIS3, ZMIZ1, LM07, RASSGR1P1 and PTPN2). In addition, at 7p12.1 COBL was the only protein-coding gene of the two genes expressed. Using the 336 islet-expressed genes as input, DAPPLE detected 123 direct interactions. A network consisting of only direct connections contained 62 genes of which nine were candidate genes (NCOA1, STAT4, ZMIZ1, ERBB3, SH2B3, LM07, SMARCE1, PTPN2 and TYK2). Based on the connectivity of each of the proteins 48 genes were highlighted as significant functional candidates including several HLA genes and five candidate genes (SH2B3, GPR183, TNFAIP3, NCOA1 and ERBB3). Conclusions: These findings provide evidence that many of the genes located in T1D risk loci are expressed at the mRNA level in human islets and interact in networks. This suggests that variations that arise to alter expression levels of these candidate genes is likely to affect network function in a way that could promote deleterious events leading to T1D. Further studies are needed to elucidate the function of the islet expressed genes in the T1D pathogenesis.
Elucidating the genetics of type 2 diabetes by integrative analysis of the genome, transcriptome and epigenome of skeletal muscle and adipose tissue samples from Finnish individuals spanning the range of glucose tolerance. J.R. Huyghe1, S.C. Parker2, M.R. Erdos3, H. Koistinen4, P.S. Chines2, H.M. Stringham5, L.J. Scott2, L. Taylor7, T. Blackwell7, H. Jiang4, C. Ma5, A.U. Jackson7, R. Wolch1, N. Narisu2, A.J. Swift5, L.L. Bonnycastle6, M.L. Sittie7, L. Kinnunen1, R.M. Watmabe5,6, T. Lakka6, M. Laakso5, J. Tuomilehto6, F.S. Collins1, M. Boehnke1. 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 3) National Institute for Health and Welfare, Helsinki, Finland; 4) Department of Preventive Medicine, University of Southern California (USC) Keck School of Medicine, Los Angeles, CA, USA; 5) Department of Physiology and Biophysics, Keck School of Medicine of USC, Los Angeles, CA, USA; 6) University of Eastern Finland, Kuopio, Finland.

Genome-wide association studies (GWAS) have identified >70 loci associated with type 2 diabetes (T2D) risk, and have significantly advanced our understanding of the pathophysiology of T2D. Yet, for most identified loci, the causal genes and functional variants remain elusive because the associated region resides in noncoding DNA. Despite the incomplete annotation of the noncoding genome, there is overwhelming evidence that GWAS loci for complex traits cluster near transcriptional regulatory elements in disease-relevant tissues. Therefore, a crucial next step in functional investigations of GWAS-identified variants is the examination of their relationship to gene expression in disease-relevant tissues, collected from subjects representing different stages of progression to T2D. To that end, as part of the Finland-United States Investigation of NIDDM Genetics (FUSION) study, we collected biopsies from two major insulin target tissues: skeletal muscle and adipose tissue. Skeletal muscle accounts for ~80% of insulin-responsive glucose uptake. Adipose tissue, while not a major site of glucose disposal, plays a key role in insulin resistance and T2D by mechanisms that remain unclear. We obtained biopsy samples from vastus lateralis skeletal muscle and abdominal subcutaneous adipose tissue from 324 clinically well-characterized Finnish individuals spanning the range of glucose tolerance: 125 normal glucose-tolerant, 72 impaired glucose-tolerant, 41 impaired fasting glucose, and 86 newly diagnosed T2D cases without antihyperglycemic medication. As of May 2013, sequencing of high quality (RIN 7.3-9.1) mRNA extracts to a median depth of 46.8 million aligned read pairs has been completed for 132 samples. RNA-seq quality control metrics and analyses of spiked-in RNA controls indicate excellent data quality. Genome-wide genotype and DNA methylation data are being generated using the Illumina Omni 2.5M-Quad SNP chip and Illumina HumanMethylation450 chip. A primary study aim entails examining the relation between expression quantitative trait loci (eQTL) and GWAS-associated SNPs for T2D and related traits. Integration with epigenetic marks will help to nominate putative causal variants and genes for functional follow-up. Further, this rich data resource will enable the study of the diverse molecular processes involved in insulin resistance which is a biomedical priority as most interventions proven to delay T2D onset act to reduce insulin resistance.

Altered expression of ARAP1, at a type 2 diabetes GWAS locus, influences insulin secretion from pancreatic beta cells. J.R. Kulzer, M.P. Fogarty, K.L. Mohlke. Department of Genetics, UNC Chapel Hill, Chapel Hill, NC.

The biological mechanisms underlying association of the ARAP1 locus with type 2 diabetes and proinsulin levels are unknown. The risk allele of index SNP rs11603334 increases transcriptional activity at an ARAP1 promoter and is associated with increased levels of ARAP1 mRNA expression in human islets. However, the functional effects of altered ARAP1 expression on insulin processing and secretion are unknown. ARAP1 is a GTPase activating protein (GAP) that catalyzes hydrolysis of GTP bound to Arf and Rho GTPases, rendering the GTPases inactive. Arf and Rho GTPases regulate Golgi transport, membrane trafficking, and actin cytoskeleton dynamics, processes that are important to insulin processing and secretion in the beta cell. We hypothesized that ARAP1 acts through one or more Arf or Rho GTPases to exert a regulatory role on insulin processing and/or secretion. In cell culture or in vitro, ARAP1 has been shown to hydrolyze GTP-bound Arf1, Arf5, Arf6, RhoA, and Cdc42. The specific Arf and Rho GTPases regulated by ARAP1 in beta cells are unknown. To investigate the effects of altered ARAP1 expression on insulin processing and secretion, we knocked-down and overexpressed ARAP1 in 832/13 and MIN6 beta cell lines and measured levels of secreted insulin. Preliminary results show a 57% decrease in glucose-stimulated insulin secretion (P = .026) upon 55% siRNA knockdown of ARAP1. Transient overexpression of wild-type ARAP1 resulted in a 70% increase in KCl-stimulated insulin secretion (P = .028). To determine which Arf and Rho GTPases are regulated by ARAP1 in beta cells, we used GTP pull-down assays to measure the levels of GTP-bound Arf1 remaining after ARAP1 overexpression. Exogenous wild-type ARAP1 decreased Arf1-GTP levels by 47%. Exogenous ARAP1 variants with either a catalytically inactive Arf GAP domain or an inactive Rho GAP domain decreased Arf1-GTP levels by ~30%. Overexpression of an ARAP1 variant with both catalytically inactive GAP domains did not decrease Arf1-GTP levels, suggesting that both domains play a role in catalyzing Arf1-GTP hydrolysis. Arf1-GTP regulates transport between the endoplasmic reticulum and the Golgi apparatus, a critical step to the efficient packaging of secretory granules. Ongoing work assesses the effects of ARAP1 overexpression on Arf5, Arf6, Cdc42, and RhoA. Our results suggest that altered expression of ARAP1 may affect Arf1 activity and interfere with proper regulation of insulin trafficking and secretion.
826T
The widely presumed type 2 diabetes causal variant, rs7903146, within TCF7L2 binds a specific protein complex not seen with its closest non-causal proxy, Q. Xia1, S. Deliadi2, C.X. Yuan3, M.E. Johnson4, S.F.A. Grant1,3,4, 1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Proteomics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA. There have been many efforts to resolve the underlying causative mechanism for a given GWAS signal. However, in the vast majority of situations, there still remains only a list of candidate variants that could represent the causal event. The situation is somewhat more advanced for the strongest associated type 2 diabetes (T2D) locus reported to date, TCF7L2, where follow-up analyses in multiple ethnicities have strongly implicated rs7903146 as the causal variant within intron 3. As such, we carried out oligo pull-down combined with mass spectrophotometry (MS) to elucidate the transcriptional machinery across the SNP. Nuclear lysates from HCT1116 cells, where TCF7L2 is abundantly expressed, were incubated with biotin-labeled, double-stranded 60bp oligonucleotides spanning rs7903146. The DNA-protein complexes were precipitated with streptavidin-agarose beads, and the bound proteins were isolated by denaturing SDS-PAGE. One major band was observed with the rs7903146-specific oligo that was absent with the scrambled oligo. Furthermore, an oligo coinciding with a SNP in strong linkage disequilibrium with rs7903146 in Caucasians but not in other ethnicities and widely rejected as the causal variant, namely the proxy rs12255372, also did not yield this extra band. Following digestion with trypsin, the samples were analyzed by MS. We observed that poly (ADP-ribose) polymerase 1 (PARP-1) is by far the most abundant binding factor. Among the next most abundant binding proteins, a number were shown to dimerize with the TCF7L2 protein, which resonates with our previous TCF7L2 ChIP-seq work. In addition, we observed differences between alleles for rs7903146. We also found evidence for an allelic difference in the MS results for proteins with less abundant binding, namely X-ray repair cross-complementing 5 (XRCC5) and RPA, which is consistent with the T2D risk. Furthermore, we implicate PARP-1 as playing a role in the T2D pathogenesis, a target which has classically been pursued for cancer.

826F
A role for IRF6 in cell cycle regulation and DNA damage response in mesenchymal stem cells. G.S. Kobayashi1, L. Alvizz1, B.V.P. Almada1, L.C. Andrade-Lima2, C.F. Menck1, M.R. Passos-Bueno1, 1) Human Genome Research Center, Institute for Biosciences, University of São Paulo, Brazil; 2) Institute of Biomedical Sciences, University of São Paulo, Brazil. Purpose: The transcription factor IRF6, which regulates orofacial and epidermal development, is associated with both skin barrier dysfunction and non-syndromic cases of cleft lip/palate (CL/P). The most prevalent syndromic form of CL/P is van der Woude syndrome (VWS), which is thought to be caused by haploinsufficiency of IRF6. Despite the established role of IRF6 in regulating cell cycle and cell cycle checkpoints in some normal-origin, little is known about its functions in non-epithelial cells. We have recently reported that mesenchymal stem cells from non-syndromic CL/P patients exhibit impairment of DNA repair mechanisms, which are known to be tightly related to cell cycle regulation (Kobayashi & Alvizzi et al., 2013). Therefore, our objective was to verify if IRF6 participates in these cellular processes in mesenchymal stem cells, in order to further clarify its biological role in non-epithelial cells and in the aetiology of CL/P itself. Methods: We quantified DNA double-strand break (DSB) formation using flow cytometry for anti-H2AX at 6 and 24 hours after exposure to H2O2 (100µM), using 2 VWS and 2 control orbicularis oris muscle-derived stem cell (OoMDSmc) cultures. The proliferative profile of the cell cultures was assessed using an XTT-based assay. Real-time quantitative PCR (RT-qPCR) was carried out in 3 independent experiments to measure IRF6 mRNA expression in untreated and G2/M-synchronized OoMDSmc (serum-starved for 48 hours and contact-inhibited under confluence). Synchronisation was confirmed by flow cytometry analysis using propidium iodide and RT-qPCR for proliferation marker MKI67. Results/Conclusions: We observed that the combination of TCF7L2 and RPA/p70. Our results point to a protein complex binding across the SNP. Nuclear lysates from HCT116 cells, where TCF7L2 is abundantly expressed, were incubated with biotin-labeled, double-stranded 60bp oligonucleotides spanning rs7903146. The DNA-protein complexes were precipitated with streptavidin-agarose beads, and the bound proteins were isolated by denaturing SDS-PAGE. One major band was observed with the rs7903146-specific oligo that was absent with the scrambled oligo. Furthermore, an oligo coinciding with a SNP in strong linkage disequilibrium with rs7903146 in Caucasians but not in other ethnicities and widely rejected as the causal variant, namely the proxy rs12255372, also did not yield this extra band. Following digestion with trypsin, the samples were analyzed by MS. We observed that poly (ADP-ribose) polymerase 1 (PARP-1) is by far the most abundant binding factor. Among the next most abundant binding proteins, a number were shown to dimerize with the TCF7L2 protein, which resonates with our previous TCF7L2 ChIP-seq work. In addition, we observed differences between alleles for rs7903146. We also found evidence for an allelic difference in the MS results for proteins with less abundant binding, namely X-ray repair cross-complementing 5 (XRCC5) and RPA, which is consistent with the T2D risk. Furthermore, we implicate PARP-1 as playing a role in the T2D pathogenesis, a target which has classically been pursued for cancer.
Meta-analysis on eQTL mapping identify insertion and deletion (INDEL) specific eQTLs in LCL, PBMC and skin tissues. J. Huang, J. Chen, J. Esparza, J. Ding, J. Elder, Y. Lee, M. Moffatt, W. Cookson, L. Liang. 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Max-Delbrück-Center for Molecular Medicine, Berlin, Germany; 4) Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 5) Department of Dermatology, University of Michigan Medical School, Ann Arbor, MI 48109-0932, USA; 6) National Heart and Lung Institute, Imperial College London, London, UK.

Large scale genome-wide association studies (GWAS) for gene expression quantitative trait loci (eQTL) mapping have been primarily focused on single nucleotide polymorphisms and have helped interpret findings from GWAS for a variety of complex diseases and traits. The functional effect of structure variants, especially short insertions and deletions (INDEL) have not been systematically investigated due to limitation in the availability of high throughput techniques. In this study, we imputed 1,380,133 INDELS based on the latest 1000Genomes project panel into 3 eQTL datasets from multiple tissues including lymphoblastoid cell lines (LCLs), peripheral blood mononuclear cell (PBMC) and skin. An eQTL Meta analysis combining 741 samples from these datasets identified INDEL specific eQTLs for 325 genes (FDR<5%), which cannot be mapped to any SNP at genome-wide significant level even after imputation. Among these significant INDEL eQTL pairs, 3,232 (1.9%) show significant tissue specific effect, compared to 9,775 (1.5%) for significant SNP eQTL pairs. We also show that INDEL specific eQTLs were enriched in Gene Ontology categories apoptosis, cell adhesion and cell cycle. Finally, imputation of 1000Genomes SNP, INDEL and cross-tissue meta-analysis together identify cis eQTLs for 6,228 genes. All eQTL results are available on our newly developed web browser. This study provides new insights into the underlying genetic architecture of gene expression, and furthermore, we illustrate how this newly developed database for INDEL specific eQTLs based on multiple tissues could help explain the association to available disease phenotypes, including asthma, atopic dermatitis and psoriasis.

Integrative Genomics Approach to Unravel the Molecular Mechanisms Underlying Genome-Wide Association Results For Lung Function Measures. M. Obeidat, K. Hao, Y. Bosse, D. Nickle, D. Postma, M. Lavoile, A. Sandford, D. Daley, C. Brandsma, M. Berge, R. Vesseley, G. Opitiec, W. Timens, D. Sin, P. Paras. 1) UBC James Hogg Research Centre, Vancouver, BC, Canada; 2) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, USA; 3) Department of Molecular Medicine, Laval University, Québec, Canada; 4) Institut universitaire de cardiologie et de pneumologie de Québec, Laval University, Québec, Canada; 5) Merck Research Laboratories, Boston, MA, USA; 6) Department of Pulmonology, University of Groningen, University Medical Center, Groningen, GRIAC research institute, Groningen, The Netherlands; 7) Department of Pathology and Medical Biology, University of Groningen, Groningen, University Medical Center Groningen, GRIAC research institute, Groningen, The Netherlands; 8) Merck & Co. Inc., Rahway, NJ, USA.

Background: SpiroMeta and CHARGE consortia published a large scale meta-analysis of lung function measures (n~49,000), which identified 26 novel loci (Artigas et al. Nat. Gen. 2011). However, the exact mechanisms underlying these associations are not fully understood. Hypothesis: A subset of SNPs which influence lung function act as eQTL in the lung to change the level of expression of their gene product. Aim: To identify SNPs that are associated with both lung function measures and mRNA levels (eQTL) in lungs. Methods: The lung eQTL were derived from genome-wide genotyping and gene expression analysis of 1,111 lung tissue samples. The study identified ~470,000 SNPs related to the level of gene expression in cis and ~17,000 SNPs in trans, at 0.1 FDR (Ke Hao et al. PLoS Gen 2012) The SpiroMeta and CHARGE consortia have made available the top 2000 SNPs associated with forced expiratory volume in one second (FEV1), and its ratio to Forced Vital Capacity (FEV1/FVC). We undertook an integration of these 4000 SNPs with the lung specific eQTL at a 0.1 FDR. Results: From the CHARGE study associations with FEV1, 1293 SNPs were associated with cis eQTL and 54 with trans eQTL, and among the FEV1/FVC associated variants, 809 were associated with cis eQTL and 236 with trans eQTL. For SpiroMeta associations’ results for FEV1, 594 SNPs were associated with cis eQTL and 54 with trans eQTL, and among the FEV1/FVC associated variants, 576 SNPs were associated with cis eQTL and 91 SNPs with trans eQTL. 48 eQTL overlapped between SpiroMeta and CHARGE for FEV1 and an equal number (48 eQTLs) for FEV1/FVC, and 7 eQTL were common to both consortia associations with FEV1 and FEV1/FVC. The mRNA levels of a number of genes regulated by the integrated SNPs also correlate with lung function measures in the lung eQTL study. Gene Ontology and pathway enrichment analyses showed lung function genes to be involved in inflammatory, tissue remodeling and developmental pathways. Future work will focus on prioritizing a number of genes for validation at protein level. Conclusion: A large number of lung function associated SNPs act as lung specific eQTL for either FEV1 or FEV1/FVC in the two consortia studied. Additionally 7 eQTL were overlapping in the two cohorts and affected both FEV1 and FEV1/FVC. The data suggest that a number of the identified variants influence lung function by modulating gene expression levels in lung.
831W

Identifying the molecular mechanisms at the vascular PLXND1 locus associated with human waist-hip ratio. T. S. Roman1, J. E. N. Minchin2, J. F. Rawls1, K. L. Mohike3, 1) Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Molecular Genetics and Microbiology, Duke University, Durham, NC.

Genome-wide association studies for waist-hip ratio (WHR), a measure of body fat distribution, identify numerous loci, including several located near genes related to angiogenesis. Stimulation of angiogenesis alters adipose metabolism, and inhibition of angiogenesis abrogates adipose tissue expansion. One locus associated with WHR by the GIANT consortium is PLXND1 (Genetics: the University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Molecular Genetics and Microbiology, Duke University, Durham, NC). PLXND1 encodes Plexin D1, a transmembrane co-receptor expressed in endothelial cells that plays a role in physiological and pathological angiogenesis. PLXnd1−/− null mice have vascular defects and die shortly after birth, and plxnd1−/− null zebrafish are viable but also have vascular patterning defects. Human adipose tissue biopsies showed a positive correlation between PLXND1 mRNA levels and a hypertrrophic adipose morphology in visceral but not subcutaneous adipose, suggesting adipose-specific effects of Plexin D1 signaling. The DNA elements regulating PLXND1 expression and the identity of the underlying functional GWAS variants are unknown. We hypothesize that WHR-associated DNA variants are located within cis-regulatory DNA elements that control tissue-specific expression of PLXND1 and regulate fat deposition and morphology. All 37 WHR-associated DNA variants at the PLXND1 locus are non-coding (r>5, 1000 Genomes Phase 1 CEU), and all overlap with open chromatin or histone modifications that mark gene regulatory regions in human umbilical vein endothelial cells (HUVEC). We defined and tested cis-regulatory elements for effects on PLXND1 expression using zebrafish transgenesis reporter assays and characterized their enrichment in a computational analysis. Candidate cis-regulatory elements recombined into a GFP expression vector were injected into one-cell stage Tg(flk1:mCherry) zebrafish embryos to test for regulatory activity. This zebrafish reporter assay enables identification of temporal-specific or cell-type specific regulatory elements. The same candidate cis-regulatory elements were tested in transcriptional reporter assays transfected into HUVEC. Preliminary data suggests that at least 3 regulatory elements show greater than 1.5-fold enhancer activity relative to an empty vector control in HUVEC. Identification of these regulatory elements at PLXND1 will provide greater insight into the molecular mechanisms and cell types important for regulating PLXND1 expression, and lead to a greater understanding of how PLXND1 influences WHR and fat deposition in humans.

831F

Enabling phenomics with high-throughput whole-organism 3D pan-cellular imaging at cell resolution. K. Cheng1, P. La Riviere2, 1) Jake Gittlen CA Res Inst, H059, Penn State College of Medicine, Hershey, PA; 2) Dept of Radiology, University of Chicago, Chicago, IL.

Genes and environment can impact any cell type at any developmental stage. Ideally, the comprehensive study of those influences in a multicellular organism requires the ability to visualize, identify, and characterize every cell type in the whole organism across all life stages at cell resolution. The most common medical implementation of cell imaging, histology, is achieved using tissue sections of about 5 micron thickness, differential color staining, and bright-field imaging at high magnification in the range of 1 micron. The destructive and tedious nature of histology precludes the study hundreds of thousands of samples, as would be required to study the functions or effects of tens of thousands of genes or chemicals. A collaborative team is working towards the ideal of high-throughput tomographic imaging and characterization of all cell types in whole small model organisms and millimeter scale tissue samples in a way that is immune to tissue opacity and pigmentation, based on the use of X-rays. Pan-cellular stains, fields of view, and phase contrast through use of monochromatic X-rays are possible at the Advanced Photon Source at Argonne National Labs and have made possible 3D imaging of whole zebrafish at larval and juvenile stages at cell resolution. We are able to recognize and characterize virtually all cell types and achieve quantitative imaging of samples containing multiple cell types. We propose that high-throughput adaptations of pan-cellular whole-organism microCT can be used as a powerful foundation for a computational phenomics that will allow us to probe the function of all genes (genetic phenomics) and the effects of environmental manipulations including chemical exposure and disease, and food treatments. The result is a high-throughput, quantitative morphometric analysis and visualization can be eliminated by adoption of existing and imminent technologies. We envision the creation of one or more beamedlines at the Advanced Photon Source at Argonne National Labs dedicated to high-throughput imaging of millimeter scale specimens, including whole invertebrate organisms such as zebrafish, 3D images can be created, analyzed and visualized, and made available to the public in real time. Such data will tremendously enhance our understanding of the impacts of genes and environment on biological systems.

832T

Genome-wide enrichments for regulatory elements across thousands of unlinked disease-associated variants. A.K. Sarkar1,2, L. Ward1, M. Kellis1,2, 1) Massachusetts Institute of Technology, Cambridge, MA, USA; 2) The Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Genome-wide association studies have identified thousands of non-coding genetic variants associated with disease, but many additional variants lie below the stringent p-value threshold given current cohort sizes. Here, we investigate both top-scoring and weakly-associated variants using functional genomics annotations of regulatory regions in 100 human tissues and cell lines from ENCODE and the Epigenomics Roadmap project. We find top-scoring regulatory regions are highly enriched for regulatory annotations of relevant cell lines, suggesting regulatory variation may be contributing to the molecular basis of the disease phenotype, and guiding the search for relevant cell types for QTL studies in a number of cases. We also find some surprising enrichments, suggesting new and unexpected tissue and cell types may be playing roles in previously unsuspected diseases, which can help guide directed experiments to reveal potentially novel pathways, tissues, and cell types. We also tested the enrichment of non-coding variants below the genome-wide significance threshold (p-value, 10−8, 1000 Genomes Phase 1 CEU) for 243 disease, quantitative trait loci (QTLs) in 5257 individuals from the Framingham Heart Study. At FDR <0.05, we identified 886 cis-eQTLs (local regulation) and 59 trans-eQTLs (distant regulation) for 243 complex traits across all life stages at cell resolution. We then explored the genetic associations of trait-associated SNPs (expression quantitative trait loci, eQTLs) in 5257 individuals from the Framingham Heart Study. At FDR <0.05, we identified 886 cis-eQTLs (local regulation) and 59 trans-eQTLs (distant regulation) for 243 complex traits and diseases. Using 4967 SNPs associated with BMI (P<5.0×10−8) with 427 complex traits from dbGaP and the NHGRI GWAS Catalog, we constructed a disease network by virtue of common SNP associations. We found several examples of genetic overlap between traits (e.g., PLXND1 in turn is also significantly associated with BMI, P=5.8×10−10), and expression of PTEN in turn is also significantly associated with BMI (P=4.5×10−10), suggesting that some genes are part of an expanding network function by altering mRNA expression. In exploring eQTLs across multiple tissues (lymphoblastoid, liver and brain) from GTEx and MRCA database, we found 40% of eQTLs on average appeared in more than one tissue, suggesting many trait-associated eQTLs found in blood may have impacts in other tissues. Our findings may provide new insight into complex genetic regulatory mechanisms.

834W

Genetic variants and regulation in human complex diseases and traits. C. Yao1, R. Joehanns2, A.D. Johnson2, B. Li2, P. La Riviere1, J.F. Rawls1, 1) Jake Gittlen CA Res Inst, H059, Penn State College of Medicine, Hershey, PA; 2) Dept of Radiology, University of Chicago, Chicago, IL.

Genetic variants and environmental factors (including nutrition) can impact any cell type at any developmental stage. Ideally, the comprehensive study of those influences in a multicellular organism requires the ability to visualize, identify, and characterize every cell type in the whole organism across all life stages at cell resolution. The most common medical implementation of cell imaging, histology, is achieved using tissue sections of about 5 micron thickness, differential color staining, and bright-field imaging at high magnification in the range of 1 micron. The destructive and tedious nature of histology precludes the study hundreds of thousands of samples, as would be required to study the functions or effects of tens of thousands of genes or chemicals. A collaborative team is working towards the ideal of high-throughput tomographic imaging and characterization of all cell types in whole small model organisms and millimeter scale tissue samples in a way that is immune to tissue opacity and pigmentation, based on the use of X-rays. Pan-cellular stains, fields of view, and phase contrast through use of monochromatic X-rays are possible at the Advanced Photon Source at Argonne National Labs and have made possible 3D imaging of whole zebrafish at larval and juvenile stages at cell resolution. We are able to recognize and characterize virtually all cell types and achieve quantitative imaging of samples containing multiple cell types. We propose that high-throughput adaptations of pan-cellular whole-organism microCT can be used as a powerful foundation for a computational phenomics that will allow us to probe the function of all genes (genetic phenomics) and the effects of environmental manipulations including chemical exposure and disease, and food treatments. The result is a high-throughput, quantitative morphometric analysis and visualization can be eliminated by adoption of existing and imminent technologies. We envision the creation of one or more beamedlines at the Advanced Photon Source at Argonne National Labs dedicated to high-throughput imaging of millimeter scale specimens, including whole invertebrate organisms such as zebrafish, 3D images can be created, analyzed and visualized, and made available to the public in real time. Such data will tremendously enhance our understanding of the impacts of genes and environment on biological systems.

Posters: Complex Traits and Polygenic Disorders
835T
READ1, a Regulatory Element within DCCD2, Epistatically Affects Reading and Language with both Deleterious and Protective Alleles. N.R. Powers1,2, J.D. Eicher1, Y. Kong1, L.L. Miller3, S.M. Ring4, J.R. Gruen1,2,3,4. 1) Dept of Genetics, Yale University, New Haven, CT; 2) Dept of Pediatrics, Yale University, New Haven, CT; 3) Dept of Investigative Medicine, Yale University, New Haven, CT; 4) Dept of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; 5) W.M. Keck Biotechnology Resource Laboratory, Yale University, New Haven, CT; 6) School of Social and Community Medicine, University of Bristol, Bristol, UK.

Learning disabilities are neurobehavioral disorders involving unexpected difficulty with reading, writing, and other aspects of language. We are currently applying this genome-wide to MS GWAS signals to identify implicated genes and will discuss these findings.

traits. We are currently applying this genome-wide to MS GWAS signals to correlate DHSs (p < 2.5e-4) comprise less than 4% of all DHSs surrounding 700 women in North America, over 60% of whom will die as a direct result. Multiple sclerosis (MS) is a progressive neurological disease affecting ~1/1,000 women in North America, over 60% of whom will die as a direct result. In the Avon Longitudinal Study of Parents and Children (ALSPAC), a large prospective birth cohort based in the UK, we showed a six-marker risk haplotype in strong linkage disequilibrium with allele 5 of READ1 to be strongly associated with severe RD. We also showed another risk haplotype in the same haplotypic block and in strong linkage disequilibrium with allele 6 of READ1 to be strongly associated with LI. We subsequently showed READ1 to bind the potent transcriptional silencer ETV6, and to participate in a synergistic genetic interaction with a known risk haplotype in the 5’ region of KIAA0319, another known RD-associated gene, to adversely affect neuromuscular disease.

835W
TNF-β Nco1 polymorphism and sepsis susceptibility following major elective surgery. R.Nath. Srivastava1, K. Baghel1,2, S. Raj1,2, A. Chandra1, S.K. Goel1. 1) Orthopaedic Surgery, King George’s Medical University, Lucknow, Uttar Pradesh, India; 2) Surgical Gastroenterology, King George’s Medical University, Lucknow, Uttar Pradesh, India; 3) Petrochemical Toxicology Division, Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India.

Background: Postoperative sepsis remains a significant cause of morbidity and mortality. In injured patients, it has been shown that a polymorphism of the tumor necrosis factor-β (TNF-β) gene is related to the development of sepsis. We investigated relation of TNF-β gene polymorphism and serum level of cytokine TNF-α with the development of sepsis after elective major surgery. Methods: The study group consists of 211 patients undergoing major elective surgery. TNF-β Nco1 polymorphism was studied in genomic DNA by analyzing restriction fragments of Nco1-digested DNA fragment using Polymerase Chain Reaction. All patients were followed for 1 month following surgery for any evidence of sepsis. Serum TNF-α levels were measured pre and postoperatively by Enzyme Linked Immunosorbent Assay. Genotypes and TNF-α production were related to the occurrence of sepsis if any. Results: 21.90% (n=46) of the patients developed postoperative sepsis. The overall mortality was 4.2% (n=9). The overall allele frequency of TNF-β genotype was 0.32 for TNF1B and 0.68 for TNF2B. In TNF-β genotype, 11.84% (n=25) patients were homozygous recessive TNF1B, 41.23% (n=87) were heterozygous TNF1B/TNF2B and 46.91% (n=99) were homozygous dominant TNF2B. Incidence of postoperative sepsis was significantly (p=0.01) higher in patients homozygous for the allele TNF2B. When compared with patients carrying at least one TNF1B allele (TNF1B homozygous and heterozygous genotype), the TNF2B homozygous genotype was associated with an Odds ratio of 2.60 (p=0.005; 95% CI 1.32 to 5.15) for the development of sepsis. Compared with the heterozygous genotype, the Odds ratio for the homozygous TNF2B genotype was 3.00 (p=0.003; 95% CI 1.39 to 6.44). In patients with postoperative sepsis, TNF-α serum cytokine levels were significantly higher (p=0.02) in TNF2B homozygous individual as compared to other genotypes. Conclusion: The development of sepsis was associated with higher capability to produce TNF-α after surgery. TNF-β Nco1 polymorphism and postoperative sepsis with increased TNF-α serum level. In patients without postoperative sepsis, the TNF-β polymorphism was not related to different levels of TNF-α production. This indicates an association between TNF-β polymorphism and postoperative sepsis, suggesting the B2/B2 genotype as a high risk factor for the development of sepsis after elective surgery.

836F
A computational framework for identifying genes perturbed by MS associated variants through regulatory element disruption. P. Shoosh-tari1,2, C. Cotsapas1,2. 1) Neurology, Yale University, New Haven, CT; 2) Broad Institute of Harvard and MIT, Cambridge, MA.

Multiple sclerosis (MS) is a progressive neurological disease affecting ~1/1,000 women in North America, over 60% of whom will die as a direct result. Genome-wide association studies (GWAS) have uncovered tens of genomic loci harboring genetic variants predisposing to the disease. As with other complex traits, most MS risk variants are likely to be due to the combined effects of many genetic variants that individually account for only a small fraction of the trait variance.

838T

The functional effects of genetic variants associated with the risk of developing rheumatoid arthritis (RA) remain elusive. To assess the impact of the RA-associated serotonin receptor 2A (HTR2A) haplotype on immune cells, we genotyped patients with established RA (n=379) for the risk haplotype via two single-nucleotide polymorphisms (SNPs) in the HTR2A locus: rs6314 and rs1328674. Low-resolution HLA-typing was also performed, and anti-citrullinated protein antibody (ACPA) levels were measured. Patients with and without the RA-associated TC haplotype were selected and T-cell and monocyte function was monitored following in vitro stimulations with staphylococcal enterotoxin B (SEB) and lipopolysaccharide (LPS) using multiparameter flow cytometry. Within the cohort, 44 patients were heterozygous for the TC haplotype (11.6%) while none were homozygous. Upon stimulation with SEB and LPS, patients produced more proinflammatory cytokines (tumor necrosis factor-alpha (TNF-α), interleukin-17 (IL-17), interferon gamma (IFN-γ), IL-2, and IL-10) and monocytes produced higher levels of TNF-α compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively).

Asthma is a chronic inflammatory condition of the lungs, characterized by excessive responsiveness of the lungs to stimuli, in the forms of infections, allergens, and environmental irritants. Currently, 22.9 million Americans suffer from asthma, and the prevalence has increased dramatically since 1980. Asthma is the leading chronic illness in U.S. children, with 6.8 million affected in 2006. Twin studies have shown that there is a genetic element to asthma susceptibility (with heritability of the condition estimated at between 0.36 and 0.77), but only a fraction of the heritability is explained with previously identified loci. In a genome wide association screen (GWAS), we identified DENND1B as a novel asthma susceptibility locus. At least two splicing isoforms of DENND1B exist, and we have attempted to examine the relative activities of these isoforms. Overexpression of the short or long isoform of DENND1B had opposite effects on the response of cells to TNFα stimulation, with the long isoform enhancing the response and the short isoform diminishing it. The long isoform shows localization to the plasma membrane, while the short isoform shows primarily cytoplasmic staining. The conserved DENN domain functions as a guanucleotide exchange factor for the small GTPase Rab35. We have demonstrated that TNFα stimulation activates Rab35 through promotion of the GTP-bound form of the protein. We are currently determining the role of Rab35 in TNFα signaling and the influence of both DENND1B isoforms on that signaling.

840W Mild deficiency of Methylenetetrahydrofolate reductase (MTHFR) increases resistance to malaria in mice, D.N. Meadows1, M. Pyck1,2, Q. Wu, S. Torre1,2, P. Gros1,2, S. Vidal1,2,5, R. Rozen1,2.

1) Department of Human Genetics, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 2) Complex Traits Group, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 3) Department of Pediatrics, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 4) Department of Biochemistry, McGill University, Montreal, Quebec, Canada; 5) Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada.

A common polymorphism (677C→T (A222V)) in methylenetetrahydrofolate reductase (MTHFR) results in a mild enzymatic deficiency and hyperhomocysteinemia. MTHFR generates 5-methyltetrahydrofolate, the primary circulating form of folate, which is utilized in homocysteine remethylation to methionine. Homozygosity for this SNP is a risk factor for neural tube defects and may also increase risk for vascular disease, pregnancy complications and cancer. Despite these selective pressures against the T allele, homozgyosity is frequent (~5-15% in many populations, with highest levels in the Mediterranean region and Hispanic populations). To determine whether mild MTHFR deficiency may have been maintained through a selective advantage against malaria, as suggested for other human mutations in the Mediterranean region, we examined Mthfr-deficient mice and MTHFR-over-expressing mice for resistance to cerebral malaria infection with Plasmodium berghei ANKA (PbA). Male mice received 106 parasites and survival was monitored over 2 weeks. Compared with wild-type littermates, Mthfr−/− mice survived infection longer (p<0.02; log-rank test), and MTHFR+/− mice died more quickly (p<0.05; log-rank test). Percent parasitemia revealed a trend (p=0.067) toward lower values in Mthfr−/− (5.5±0.5) compared with Mthfr+/− (7.0±0.4) mice, with no differences in this parameter in MTHFR+/− mice. Splenocytes analyzed by flow cytometry with several markers for T and NK cell populations, showed increased numbers of total lymphocytes, as well as more CD4+ and CD8+ T cells and more CCR4+ NK cells in Mthfr−/− animals, compared with Mthfr+/− mice (p<0.05 for all cell types). MTHFR+/− mice showed decreased numbers of NK cells (p<0.05 as well as decreased numbers of CCR4+ NK cells (p<0.01). Serum interferon-γ (IFNγ) levels, measured by ELISA, were lower in Mthfr−/− mice (88.7±27.3 pg/mL) compared with Mthfr+/− mice (217.1±49.6 pg/mL) (p<0.05). IFNγ levels in tissues, measured by immunoblotting, were higher (by ~40%) in spleen (p<0.01) and brain (p<0.001) of Mthfr−/− compared with Mthfr+/− mice. Our results suggest that differences in the balance between pro- and anti-inflammatory immune modulators in Mthfr−/− mice may protect them from PbA infection. This is the first time that mild MTHFR deficiency has been shown to confer an advantage of malaria resistance. Understanding the mechanism underlying this unique advantage may have implications for malaria control strategies and other infections where homocysteine metabolism is involved.
843W
Identification of a regulatory variant that binds a transcriptional activator complex including FOXA1 and FOXA2 at the CDC123/CAMKD2 type 2 diabetes GWAS locus. M.P. Fogarty1, S. Vadlamudi1, M.E. Cannon1, K.J. Gaunt2, K.L. Mohlke3, 1) Dept Genetics, Univ North Carolina, Chapel Hill, NC, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

For many of the type 2 diabetes (T2D) loci identified through genome-wide association studies, signals are localized to non-protein-coding intronic and intergenic regions and likely contain variants that regulate gene transcription. The CDC123/CAMKD2 T2D association signal on chromosome 10 spans regions between rs11257655 and CAMKD2 and overlaps the CDC123 3’UTR. To gain insight into the molecular mechanisms underlying this association signal, we examined all SNPs in LD r² > 0.7 (CEU, 1000G phase 1) with the GWAS index SNP rs12775790 overlapping a predicted regulatory motif. Maps of islet and liver cell chromatin histone modifications and transcription factor ChIP-seq were used to identify potential regulatory elements. Based on these criteria, two regions containing T2D-associated variants were tested for enhancer activity using luciferase reporter assays. A 151 bp region surrounding rs11257655 located 15 kb from the 3′ end of CDC123 and 84 kb from the 5′ end of CAMKD2 displayed allele-specific transcriptional activity in rat 832/13 and mouse MIN6 insulinoma cells as well as in human HepG2 hepatocellular carcinoma cells. The rs11257655 risk allele T showed greater transcriptional activity than the non-risk allele C in all cells tested (832/13 P = 6 x 10^{-3}; MIN6 P = 2 x 10^{-5}, HepG2 P = 8 x 10^{-4}). Site-directed mutagenesis of a second included in the 151 bp region due to proximity verified rs11257655 as the driver of allelic differences in enhancer activity. Increased transcriptional activity with the rs11257655 risk allele T was consistent with the direction of an eQTL for CAMKD1 in blood (Voight, Nat Genet 42; 579). Data from the ENCODE consortium demonstrate binding of FOXA1 and FOXA2 transcription factors in HepG2 cells to a region that overlaps rs11257655. We next used site-directed mutagenesis and electrophoretic mobility shift assays to assess whether a 20 bp region containing alleles of rs11257655 differentially binds these factors. The rs11257655 T allele showed allele specific binding of FOXA1/FOXA2 in beta cell lines and hepatocellular carcinoma cells. Taken together, these results demonstrate the importance of cis-regulatory elements to activity of a complex regulatory variant that binds transcriptional factors in HepG2 cells. The rs11257655 variant is associated with decreased risk of T2D and is located in a region where risk and non-risk alleles may differentially influence enhancer function.

844T
Gene-gene interaction and RNA splicing profiles of MAP2K4 gene in rheumatoid arthritis. K. Schetinsky1, M. Ronningen1, D. Protsyuk1, L. Klare1, K. Shchetynsky2, M. Ronninger1, D. Protsyuk2, L. Klare2, R. Skog1, L. Padyukov1. 1Rheumatology Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.

Background and objectives: MAP2K4 encodes a mitogen activated protein kinase kinase 4 (MKK4), important for optimal activation of JNK1-3 and p38 - the two members of the MAP kinase family. In this study we explore the interaction between MAP2K4 locus and two major known genetic risk factors for autoantibody positive rheumatoid arthritis (RA) – HLA-DRB1 shared epitope alleles and PTPN22 rs2476601. We also address the balance in the expression of alternative MAP2K4 splice forms and its association with known SE and PTPN22 genotypes and the autoantibody profile of the disease.

Methods: The genotypes from 1985 patients with RA and 2252 matched healthy controls from the Swedish EIRA study population and from 863 RA cases and 1181 controls from the NARAC were used in the study. The interaction analysis was performed on 22 SNPs from the MAP2K4 locus, HLA-DRB1 shared epitope alleles and rs2476601 from PTPN22 by calculation the attributable proportion due to interaction (AP).

We studied transcript diversity of MAP2K4 and investigated relative expression of MAP2K4 forms in peripheral mononuclear cells for 44 RA patients and 44 controls of Caucasian ancestry. These results were analyzed against available genotypic and phenotypic data. Results: We found MAP2K4 rs10468473 in statistical interaction with SE, and PTPN22 rs2476601 in autoantibody positive RA (AP 0.197 [95% CI 0.098 - 0.296] and AP 0.28 [95% CI 0.074-0.48] respectively; EIRA+NARAC) in two independent cohorts and a combined meta-cohort. In our assessment of allelic expression, RA patients heterozygous for rs10468473 demonstrated significantly elevated MAP2K4 expression in comparison to individuals homozygous for G allele. We also observed a novel ’skipped exon’ type RNA splice variant of MAP2K4 in our study material, and a potential protein isoform of corresponding molecular weight. MAP2K4 splice forms were differentially expressed in peripheral blood mononuclear cells from 88 RA cases and controls. Within the group of RA patients, a correlation was observed between MAP2K4 variants expression and number of carried SE alleles, as well as with other phenotypic data. Conclusion: Our data suggest interaction between MAP2K4, PTPN22 and HLA-DRB1 in development of autoantibody positive rheumatoid arthritis. We also found an allele-specific effect of rs10468473 on MAP2K4 expression, and splicing events in transcripts from the MAP2K4 locus that could be relevant in disease pathogenesis.

845F
A polymorphism in human estrogen-related receptor beta (ESRRβ) associated with physiologic measures of noise-induced hearing loss. i. Bhatt1, S. Phillips1, S. Richner1, R. Grechouse1, D. Tucker1, K. Lundgren1, V. Henrik1. 1) Dept Genetics, University of North Carolina at Greensboro, Greensboro, NC, 27402, USA; 2) Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC, 27402, USA; 3) Department of Communication Sciences & Disorders, Appalachian State University, Boone, NC, 28608, USA; 4) Center for Biotechnology, Genomics & Health Research, The University of North Carolina at Greensboro, Greensboro, NC, 27402, USA.

Noise-induced hearing loss (NIHL) is a common form of hearing loss and a growing health concern despite national standards for hearing protection and public health awareness campaigns. An NIHL gene association study with college-aged musicians has associated a non-synonymous single nucleotide polymorphism (rs16742642, C→T, P386S) in the ligand-binding domain of human estrogen-related receptor beta (ESRRβ), OR = 2.8, CI = 1.4-5.9, p = 0.003 with increased susceptibility to bilateral 4 to 6 kHz hearing loss. ESRRβ is expressed in major cochlear structures except hair cells. Methods: The purpose of the study was to examine the effects of the ESRRβ polymorphism on both pre-noise exposure (baseline) cochlear physiology and on temporary changes in hearing sensitivity following a brief noise exposure. We examined baseline cochlear physiology by pure-tone audiometry, distortion product otoacoustic emissions (DPOAE growth function and high resolution DPOgram), transient-evoked otoacoustic emissions (TEOAE) and TEOAE suppression in 40 and 20 participants carrying the ESRRβ rs61742642 CC vs. CT genotype respectively. Temporary changes in hearing sensitivity were induced by 10 minutes exposure to 90 dB SL narrow-band noise (center frequency = 2 kHz) and the changes were evaluated by audiometry (at 2, 3 and 4 kHz). Results: Baseline DPOAE amplitudes were lower across the 2 to 4 kHz frequency range (DPOgram overall mean difference (MD) = 2.66 dB, p = 0.03; DP growth function overall MD = 3.30 dB, p = 0.046). Baseline TEOAE amplitudes were lower across the 2 to 4 kHz frequency range (MD = 0.35 dB, CI = 0.69 to -0.006, p = 0.046) in participants with the CT genotype. Regression analysis showed that individuals carrying the CT genotype demonstrated a greater increase in hearing loss following 10 minutes noise exposure compared to the participants with the CT genotype. Conclusions: The results indicate that individuals with the ESRRβ polymorphism show reduced outer hair cell activity possibly due to pre-chemotical degeneration. They also showed impaired activity of medial olivocochlear nerve fibers and acquired greater deterioration in hearing sensitivity following the noise exposure. These data suggest that the ESRRβ polymorphism increases susceptibility to NIHL, and also indicate the efficacy of otoacoustic emissions testing for identifying sound processing endophenotypes.
846W Genetic risk variants for autoimmune diseases regulate gene expression in thymus. J. S. M. Gabrielsen1, S. Svanström Amundsen1, H. Helgeland1, S. Tenebebe Filam1, N. Hatinoor1, K. Holm2, M. K. Viken1, B. A. Lie1, 1) Medical Genetics, Oslo University Hospital, Oslo, Oslo, Norway; 2) Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Oslo, Norway.

Genome-wide association studies (GWAS) have boosted our knowledge of genetic risk variants in autoimmune diseases (AIDs). Despite the fact that AIDs are extremely heterogeneous, GWAS indicate that certain loci and genes seem to predispose to multiple immune-related diseases. Most of the genes identified for AIDs are involved in immunological functions, but because the majority is found to be non-coding, it’s considerable to believe that these risk variants have a regulatory role. Expression quantitative trait locus (eQTL) screening is a widely used method to investigate gene regulation, and by correlating genetic association data with gene expression data, genetic risk factors acting through regulatory variants could be determined. Based on this, and the fact that thymus is an immunological important organ, we have performed an eQTL screen to investigate whether single nucleotide polymorphisms (SNPs) associated with AIDs influence gene expression in thymic tissue (i.e. whether disease associated SNPs serve cis-acting regulatory functions). The genotypes from 42 Norwegian thymic tissue samples genotyped using the Immunochip were analyzed for 420 AID-associated SNPs identified from the National Institute of Health’s catalog of GWAS (NHGRI). Gene expression was measured using Illumina Human WG-6 v3 and genotypes were correlated with gene expression of surrounding genes (+1 MB of the SNP). By logistic regression analysis, we robustly identified 7 genes (FCRL5, LOC339804, ERAF1, RNAK2, LOC100559733, LOC566222, and LOC2888814) associated with single nucleotide polymorphisms (SNPs) at a study-wide level of significance (P < 2.10^{-10}). Fine mapping of these regions led to the discovery of even stronger correlating SNPs within four regions: rs10181042 (LOC339804, P < 3.5x10^{-10}), rs30377 (ERAFA1, P < 3.3x10^{-10}), rs2293203 (RNAK2, P < 2.2x10^{-09}) and rs9917744 (LOC388814, P < 7.2x10^{-07}). The regulatory function of the risk polymorphisms and proxy SNPs in strong linkage disequilibrium (r2=0.8) were evaluated using the RegulomDB. Newly defined SNPs associated with AIDs from recent Immunochip projects were also included. Combining these resources, we have highlighted functional variants for 7 genetic regions that potentially can represent causal autoimmune risk variants. We conclude that our study shows that several autoimmune risk variants act as eQTLs in thymus.

846F Comprehensive functional assays for variants of unknown significance. J. O. Kitzmann1, L. Starita1,2, R. Lo1,2, R. Qi1, S. Field1,2, J. Shendure1, 1) Dept. Genome Sciences, Univ Washington, Seattle, WA; 2) Howard Hughes Medical Institute.

With the relative ease of variation discovery by massively parallel sequencing, functional interpretation becomes an increasingly critical bottleneck in human genetics. As a case in point, clinical diagnostics frequently reveal missense variants of uncertain significance (VUS), even in genes with clear disease relevance such as the tumor suppressor BRCA1, the primary risk factor for inherited breast cancer. To address this bottleneck, we demonstrate a massively parallel mutagenesis approach called PALS (Programmed ALlelic Series). This approach uses libraries of microarrays-derived mutagenesis primers to direct synthesis of every possible missense mutant of a gene of interest. As a proof of principle, we designed a library comprising all possible single-codon substitutions in the human tumor suppressor gene TP53 and a similar library for the well-characterized yeast transcription factor GAL4. Deep sequencing and tag-directed assembly (Hiatt et al., Nat Methods 2010) revealed that each library covered the majority of the designed mutational space. To demonstrate how PALS libraries could be used to comprehensively profile loss-of-function mutations, we subjected the GAL4 library to a function-dependent, in vivo screen in budding yeast. In this type of selection experiment, mutant haplotypes’ fitness values are determined by deep sequencing to estimate their abundances before and after selection. This assay revealed a subset of missense variants which, together with the clear loss-of-function variants (premature stops and frameshift mutations), exhibited no activity and were depleted by selection. This approach - massively parallel synthesis and sequencing coupled to functional selection - provides a general framework for deep surveys of the functional impacts of mutations to clinically relevant genes.

847W Expression of Two Genes from the Candidate Locus of Chromosome X in Rheumatoid Arthritis and Systemic Sclerosis. L. M. Diaz-Gallo1,2, K. Shcheytinsky1, A. Nordin1, A.I. Catrina1, J. Martin1,2, L. Padyukov1, 1) Karolinska Institutet, Stockholm, Stockholm, Sweden; 2) Udi para de Patologia y Biomedicina Lopez-Neyra, IPhLN-CSIC, Granada, Spain.

There is increasing evidence of association between variants at Xq28 genomic region and autoimmune diseases (AIDs). The genes interleukin-1 receptor-associated kinase 1 (IRAK1) and methyl-CpG-binding protein 2 (MECP2) are located there. We evaluated if there was differential mRNA expression of the IRAK1 and MECP2 between rheumatoid arthritis (RA) patients, systemic sclerosis (SSc) patients and healthy controls. Additionally, we assessed the correlation of IRAK1 and MECP2 expression levels with the rs1059702 and rs17435 polymorphisms located in this locus and previous associated to AIDs. The gene expression of IRAK1 and MECP2 was measured using quantitative PCR by TaqMan assays. We studied 50 RA patients, 35 SSc patients and 52 healthy controls from a European descendent population, all females. Prespecified TaqMan 5’ SNP genotyping assays were used to genotype the two SNPs. The RNA was extracted from the studied subjects. The IRAK1 and MECP2 expression was significantly different between the RA patients, SSc patients and controls (p=0.0003, p=0.0006 respectively, Kruskal-Wallis test). When we compared each group of patients against controls we observed that the IRAK1 expression was significantly decreased in the RA patients compared with controls (p=0.0022; RA relative quantity (RQ) mean=42.37; controls RQmean=60.28, Mann-Whitney test). But there was no significant difference between the levels of IRAK1 expression between SSc patients and controls (p=0.39; SSc RQmean=46.8; controls RQmean=42.12). We found no significant difference in the MECP2 expression levels between RA patients and controls (p=0.05; RA RQmean=45.72; controls RQmean=57.06). Meanwhile, the aforementioned pathogenesis was not statistically significant in the MECP2 expression between RA patients and controls with p=0.017. SSc RQmean=51.86; controls RQmean=38.71. Finally, we observed that there was a moderated positive correlation between the IRAK1 and MECP2 expression in the studied individuals (correlation coefficient=0.613, p<0.0001). Although the difference was not statistically significant, the homozogous for the minor allele of the rs1059102 trend to express higher levels of both IRAK1 and MECP2 genes. Our study showed an altered expression of IRAK1 and MECP2 genes in both RA and SSc compared to the controls. This study increases the evidence that suggest an important pathogenic role of this X chromosome locus in AIDs.

847T The apical sodium-dependent bile acid transporter ASBT (SLC10A2) affects the progression of primary biliary cirrhosis in Japanese patients via its transcriptional activity. K. Taira1, T. Inamine1, A. Kawachiuchi1, S. Kondo1, M. Nakamura2, K. Tsukamoto1, 1) Dept Pharmacothequpeutics, Nagasaki Univ Grad Sch, Nagasaki, Japan; 2) Dept Hepatology, Nagasaki Univ Grad Sch, Omura, Japan; 3) Clin Res Center, Natl Hosp Organi Nagasaki Med Center, Omura, Japan.

PURPOSE Primary biliary cirrhosis (PBC) is a chronic liver disease characterized by destruction of the intrahepatic small bile ducts, leading to cholestasis, fibrosis, cirrhosis, and eventually liver failure. In order to identify the genetic determinants of PBC progression, we focused on apical sodium-dependent bile acid transporter (ASBT), which plays a key role in reabsorption of luminal bile acids in epithelial cells of the intestines under the enterohepatic circulation, examined an association between SLC10A2 polymorphisms and the susceptibility to PBC progression, and investigated the function of these polymorphisms. METHODS A total of 309 Japanese PBC patients were classified into the following two groups (early stage and late stage) based on liver biopsy results and/or clinical manifestations. The 12 tag single nucleotide polymorphisms (SNPs) in SLC10A2 were genotyped by PCR-restriction fragment length polymorphism or -direct DNA sequencing. Furthermore, the 2.5-kb upstream promoter sequences of SLC10A2 were subject to PCR-direct DNA sequencing in order to identify new SNPs within this region. The frequencies of alleles and genotypes of 15 SNPs including 3 newly identified SNPs were compared between 151 PBC patients in early and late stages by chi-square test. To investigate the function of the SNP associated with BPC progression, the 50-bp promoter sequence containing the allele of the associated SNP, which was incorporated into luciferase reporter plasmid, was transfected into Caco-2 cells, and luciferase activities were compared between the two groups in the associated SNP. RESULTS Five SNPs of SLC10A2, including 3 newly identified SNPs in the promoter region, showed the significant association with PBC progression. Among them, one of the newly identified SNPs indicated decreased expression with PBC progression. The transcriptional expression of SLC10A2 in Caco-2 cells. CONCLUSION Our results suggest that PBC patients with the low-risk allele of SLC10A2 may represent lower expression of ASBT and then diminution of reabsorption of luminal bile acids in the intestines leading to the decrease in enterohepatic circulation of bile acids and eventually leading to slower PBC progression. Thus, SLC10A2 appears to be a genetic determinant of PBC progression in Japanese patients.
850T
An enhancer element harboring variants associated with systemic lupus erythematosus engages the TNFAIP3 promoter to influence A20 expression. S. Wang1, F. Wen1, G.E. B. Wiley1, M.T. Kinter2, P.M. Gaffney1. 1) Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma city, OK, 73104; 2) Free Radical Biology and Aging Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, 73104.

Functional characterization of causal variants present on risk haplotypes identified through genome-wide association studies (GWAS) is a primary objective of human genetics. In this report, we evaluate the function of a pair of tandem polymorphic dinucleotides, (rs148314165, rs200820567, collectively referred to as TT>A) recently nominated as causal variants responsible for genetic association of systemic lupus erythematosus (SLE) with tumor necrosis factor alpha inducible protein 3 (TNFAIP3). TNFAIP3 encodes the ubiquitin-editing enzyme, A20, a key negative regulator of NF-kB signaling. A20 expression is reduced in subjects carrying the TT>A risk alleles, however, the underlying functional mechanism by which this occurs is unclear. We used a combination of electrophoretic mobility shift assays (EMSA), mass spectrometry (MS), reporter assays, chromatin immunoprecipitation-PCR (ChIP-PCR) and chromosome conformation capture (3C) to evaluate the regions.

EBV transformed lymphoblastoid cell lines (LCL) from individuals carrying risk and non-risk TNFAIP3 haplotypes to characterize the effect of TT>A on A20 expression. Our results demonstrate that the TT>A variants reside in an enhancer element that binds NF-kB and SATB1 enabling physical interaction of the enhancer with the TNFAIP3 promoter through long-range DNA looping. Impaired binding of NF-kB to the TT>A risk alleles or knockdown of SATB1 expression by shRNA, inhibits the looping interaction resulting in reduced A20 expression. Together, these data reveal novel mechanism of TNFAIP3 transcriptional regulation and establish the functional basis by which the TT>A risk variants attenuate A20 expression through inefficient delivery of NF-kB to the TNFAIP3 promoter. These results provide critical functional evidence supporting a direct causal role for TT>A in the genetic predisposition to SLE.

851F
Gene-gene interactions of EG-VEGF, PKR1 and PKR2 genes and the risk of recurrent miscarriages. M. Su1, P. Kuo1, S. Lin2. 1) National Cheng-Kung University Hospital, Tainan City, Taiwan; 2) Institute of clinical medicine, National Cheng-Kung University, Tainan City, Taiwan.

Background: Endocrine gland-derived vascular endothelial growth factor (EG-VEGF) and its receptor genes (PKR1 and PKR2) play an important role in human early pregnancy. Our previous study showed that PKR1 and PKR2 polymorphisms are associated with recurrent miscarriages (RM). This study was conducted to find EG-VEGF, PKR1 and PKR2 variants in the coding regions of idiopathic RPL patients and further evaluate gene-gene interactions in 3 genes. Methods: Two hundred and ninety one blood samples from 142 RPL women and 149 controls were nucleotide sequenced in the coding regions of EG-VEGF, PKR1 and PKR2. Gene-gene interaction was evaluated in 3 gene variants using multifactor dimensionality reduction (MDR) method. Result(s): One each nonsynonymous variant of 3 genes were identified, and PKR1(I379V) and PKR2(V331M) were significantly associated with idiopathic RM (p=0.006 and p=0.002, respectively). Genetic interactions were found not only between PKR1(I379V) and PKR2(V331M), but also among EG-VEGF (V67I), PKR1(I379V) and PKR2(V331M) (p<0.01 and p=0.01, respectively). Women carried low-risk genotypes reduced 77% risk of experiencing miscarriages compared with those carried high-risk genotypes. Conclusion(s): The present study corroborates the clinical relevance of the EG-VEGF system in human early pregnancy, and provides evidence for the gene-gene interactions of EG-VEGF and PKR variants.

852W
Gene-asbestos exposure interactions on lung cancer risk. C. Liu1,2, 1. Stucker3, C. Chen4,5,5, G. Goodman1, M.K. McHugh2, A.M. D’Amelio8,9, C.J. Etzel4, S. Li1, X. Lin5, D.C. Christiani1,2, 1) Dept Environmental Health, HSPH, Boston, MA, USA; 2) Institute of Environmental Health, College of Public Health, National Taiwan University, Taipei, Taiwan; 3) INSERM U 754-IFR69, Villejuif, France; 4) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Mailstop M5-C800, Seattle, WA 98109-1024, USA; 5) Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA; 6) Department of Otologyngology; Head and Neck Surgery, School of Medicine, University of Washington, Seattle, WA, USA; 7) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Mailstop M3-A306, Seattle, WA 98109-1024, USA; 8) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, 1155 Pressler Boulevard, Unit 1340, Houston, TX 77030, USA; 9) Biomath/Biostatistics Program, The University of Texas Graduate School for Biomedical Sciences, Houston, TX 77030, USA; 10) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 11) Department of Medicine, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA.

Occupational asbestos exposure has been found to increase lung cancer risk based on epidemiologic studies. In order to clarify the respective roles of genetic factors and asbestos exposure on lung cancer risk, we conducted the asbestos exposure-gene interaction analyses among Caucasian populations who are current or ex-smokers. The stage 1 discovery stage included 833 Caucasian cases and 739 Caucasian controls using Illumina Human 610-Quad BeadChips. Several independent populations were included in the second replication stage. Cumulative lifetime asbestos exposure score (AES), which was previously developed and validated in our study population, was calculated from self-reported duration and intensity of occupational and nonoccupational exposures. The top ranked SNPs from discovery stage were replicated within International Lung and Cancer Consortium (ILCCO). A two-stage replication approach was conducted. First in silico replication was conducted in those groups that have GWAS and asbestos exposure data, including 1548 cases and 1544 controls. Followed by de novo genotyping to replicate the results from in silico replication, 1539 cases and 1761 controls were genotyped. In order to have the data comparable between studies, genotyping results were imputed by MACH using 1000 Genome dataset. Asbestos exposure measurements were recategorized as 'high' or 'low and none' asbestos exposure effects based on each group’s previous findings. Logistic regression adjusted for potential confounders was used to assess the SNP-asbestos exposure interaction effect on lung cancer risk. The top associated findings were found at 22q13.31, 8p22, 16p11.2, which have been reported in association with carcinogenesis and inflammatory response. Further functional analysis will be conducted among the regions.
853T

**RBMS1 genotype strongly influences adiposity and liver function in rural but not urban dwellers. M.E.S. Bailey¹, C.A. Cells-Morales², M. Staunton¹, N. Ulloa³, C. Calvo³, F. Perez-Bravo³, J.M.R. Gill³. ¹School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom; ²Human Nutrition Research Centre, Institute for Ageing and Health, Newcastle University, Newcastle, UK; ³Department of Clinical Biochemistry and Immunology, Faculty of Pharmacy. University of Concepcion, Chile; 4) Laboratory of Nutritional Genomics, Department of Nutrition, Faculty of Medicine, University of Chile, Santiago, Chile; 5) Institute of Cardiovascular and Medical Sciences, CMVLS, University of Glasgow, Glasgow, Scotland.**

**RBMS1 encodes a set of single-stranded DNA-binding proteins that may have roles in regulation of DNA replication, gene expression, cell cycle and cell fate. GWAS meta-analysis and other studies have implicated RBMS1 as a risk gene in obesity and type 2 diabetes. We show that RBMS1 strongly influences levels of subcutaneous body fat and cardiometabolic risk-related phenotypes across populations but only in those living a more traditional lifestyle. We analysed an RBMS1 SNP, rs7593730 (C/T alleles), for association with a range of quantitative trait measures in a cross-sectional population sample from Chile (n=300), including male and female adults of both native American (Mapuche) and European ethnicity living in urban and rural environments. Participants were assessed for a wide range of physical, metabolic, social and environmental variables, including body fat (aggregate of 4 skinfold measures). The T allele (p=0.14 in Europeans, 0.05 in Mapuche) was found to be strongly associated with lower levels of subcutaneous fat, insulin resistance (HOMA-IR) and ALT, and with higher levels of HDL cholesterol (all p<0.00015), with similar effect sizes in Europeans and Mapuches. The association with body fat was still preserved (genotype p=0.009) after adjustment for age, sex, ethnicity, sociodemographic factors, activity (sedentary time, MVPA) and overall food intake. This SNP was also weakly associated with several other variables, including cardiore respiratory fitness. The genotypic association with HOMA-IR was partially independent (genotype p=0.009) of body fat. Genotype was found, in fully adjusted models, to interact strongly with both living environment (rural vs urban; interaction p=0.0004) and fitness (interaction p=0.006) in its effect on adiposity. In stratified analyses, genotype was found to influence adiposity levels in rural dwellers but not urban dwellers, and more strongly in fitter individuals. In fully adjusted models, rs7593730 explained >17% of the variance in adiposity in the rural group (p=0.00001), each copy of the T allele lowering fat levels by 0.22 SDs. These findings support the idea that there may be subgroups defined by genotype in whom activity/boosting interventions to reduce adiposity may be more effective. Furthermore, they suggest that a portion of the cardiometabolic disease risk associated with a Westernised lifestyle may result from the suppression of specific genetic influences on primary risk factors in the urban environment.

855W

**Gene-carbohydrate and gene fiber interactions and type 2 diabetes in diverse populations from the National Health and Examination Surveys as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study. R. Villegas, R.J. Goodloe, B. McClellan Jr, J. Boston, D.C. Crawford. Med, Vanderbilt Univ, Nashville, TN.**

Both environmental and genetic factors impact Type 2 diabetes (T2D). Environmental modifiers of known genotype-T2D associations may account for some of the "missing heritability" of these traits. To identify such modifiers, we genotyped 15 T2D-associated variants identified through genome-wide association studies (GWAS) in 6414 non-Hispanic white, 3037 non-Hispanic black and 3633 Mexican American samples collected for the National Health and Nutrition Examination Surveys (NHANES). In this paper we evaluated interactions between these variants with carbohydrate intake and with fiber intake. We performed logistic regression analysis testing for SNP associations with T2D levels in self-identified European American, African American and Mexican American. Only 1 association generalized across all three populations, rs7903146 (TCF7L2). High intake of carbohydrates was associated with higher risk of T2D while fiber was inversely related to T2D risk in analysis conducted in all participants and in analysis stratified by race. However the trend for the association between fiber and T2D in blacks was no significant and the association between carbohydrate intake and T2D in Mexicans was of marginal significance. We identified one SNP x carbohydrates interactions at a significance threshold of p < 0.05, in blacks (with IGFBP2, rs4402960) and one in Mexican Americans (with PPARG rs1801282). We found one gene-fiber interactions with ADAMT59 rs4607103 in African Americans and another gene-fiber interaction among Mexican Americans with PPARG rs1801282, both at P<0.05 level. We also evaluated two gene-carbohydrate interactions with RLBP1 in Europeans and PPARG in African Americans and Mexican Americans. Our results suggest that carbohydrate and fiber may modify genotype-phenotype associations and that this association differs by race.

858F

**Gene-Lifestyle Interaction and Type 2 Diabetes. R.A Scott on behalf of The InterAct Consortium. MRC Epidemiology Unit, Cambridge, United Kingdom.**

Rapid progress has been made in the understanding of the genetic basis of type 2 diabetes (T2D). It is uncertain whether testing for genetic susceptibility to T2D may be useful to guide decisions about lifestyle interventions. The InterAct study includes 12,403 incident T2D cases and a representative subsample of 16,154 individuals from a total cohort of 340,234 participants from 8 European countries followed for 3.99 million person-years. We studied the combined effects of an additive genetic T2D risk score comprising 49 variants associated with T2D and mifiable and non-modifiable risk factors using Prentice-weighted Cox-regression and random effects meta-analysis methods. The effect of the genetic score differed significantly by age at study entry (p<0.01), due to a larger genetic effect in cases who developed diabetes early (<55 years), compared to later. Relative genetic risk (per standard deviation (4.4 risk alleles)) was also larger in participants who were leaner, both in terms of BMI and waist circumference (WC) (HR normal weight 1.52 (1.50, 1.54), overweight 1.46 (1.37, 1.56), obese 1.27 (1.17, 1.39), p<7.5x10⁻³³). HR low WC 1.60 (1.49, 1.72), medium WC 1.53 (1.39, 1.68), large WC 1.29 (1.18, 1.40), p<7.4x10⁻³⁰. Only one variant (ADCYS rs11717195 for SNP by BMI interaction) was below the Bonferroni-corrected threshold for individual significance (p<7.2x10⁻¹⁰), again showing a larger genetic effect size in smaller individuals. Absolute risk demonstrated the strong overall preponderance of obesity for T2D risk; the cumulative T2D incidence over 10 years (per hundred) rose from 0.25, 0.44, and 0.53 to 0.89 across quartiles of the genetic score in normal weight individuals compared to 4.22, 5.78, 5.83 and 7.99 in obese individuals. We detected no significant interactions between the genetic score and sex, diabetes family history, physical activity, or dietary habits assessed by a Mediterranean diet pattern score. The high absolute risk associated with obesity at any level of genetic risk highlights the importance of universal approaches to lifestyle intervention.
856T
Variants in the Glucagon Gene are Associated with Baseline Weight and Glycemic Response to Metformin and Intensive Lifestyle Interventions in the Diabetes Prevention Program. A.H. Winters1, K.A. Jablonski2, S.E. Kahn3, W.C. Knowler4, E.S. Horton5, K.J. Mathers6, R.F. Arakaki7, J.C. Flores8,9,10, T.I. Pollin11, DPP Research Group. 1) Human Genetics, University of Maryland, Baltimore, Baltimore, MD; 2) The George Washington University Department of Epidemiology and Biostatistics. The Biostatistics Center Rockville, MD; 3) Division of Metabolism, Endocrinology and Nutrition Department of Medicine VA Puget Sound Health Care System and University of Washington Seattle, WA; 4) Diabetes Epidemiology and Clinical Research Section National Institute of Diabetes and Digestive and Kidney Diseases Phoenix, AZ; 5) Joslin Diabetes Center, Harvard Medical School, Boston, MA; 6) Division of Endocrinology and Metabolism, Indiana University School of Medicine, Indianapolis, Indiana; 7) Department of Medicine Clinical Research, University of Hawaii, Honolulu, Hawaii; 8) Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit) Massachusetts General Hospital; 9) Program in Medical and Population Genetics Broad Institute; 10) Department of Medicine Harvard Medical School; 11) Endocrinology, Diabetes and Nutrition, University of Maryland, Baltimore, Baltimore, MD.

The glucagon gene (GCG) codes for glucagon, glucagon-like peptides 1 and 2 (GLP1 and GLP2), glicentin, and oxyntomodulin, which have a number of different physiological effects including glucose homeostasis, weight gain and satiety. We hypothesized that variation in GCG affects glycemic traits and weight and modifies glycemic response to weight- and glucose-lowering interventions. The Diabetes Prevention Program (DPP) was a clinical trial testing metformin or a lifestyle intervention (systematic caloric and fat intake reduction and increase in physical activity) versus placebo on reducing type 2 diabetes development in those at high risk. We evaluated whether SNPs in GCG modify effects of these interventions in the DPP. Thirteen SNPs tagging common variations in GCG were evaluated for association with weight, fasting and two hour glucose, glycated hemoglobin and oral disposition index at baseline and 1 year. Among 2815 DPP participants, the T allele of rs2892827 (DPP minor allele frequency=0.14-0.19) was associated with greater weight at baseline (β=1.46[0.19, 2.73] kg/T allele; p=0.02). There were no significant SNP x treatment interactions on baseline-adjusted one year weight (interaction p=0.25 for lifestyle and p = 0.24 for metformin). There was a significant interaction on fasting glucose with both treatments (p=0.02 for SNP x lifestyle and p=0.04 for SNP x metformin), explained by a borderline-adjusted fasting glucose increase of 0.15 mg/dl in the placebo group (β=1.53[0.03, 3.08]; p=0.055), but not in the lifestyle (β=−0.93[−2.14, 0.28]; p=0.13) or metformin (β=−0.59[−1.75, 0.57]; p=0.32) groups. Three other SNPs in GCG had significant interactions with lifestyle on weight change as measured by baseline-adjusted one year weight (rs11897425 p=0.03; rs13010545 p=0.006; rs13020420 p=0.02), showing evidence that lifestyle reduces the effect of the SNP on weight. No other significant interactions with metformin were seen. rs2892827 is 5’ of GCG near its known transcription factor binding sites, and the major A allele is conserved in mammals. Thus the T allele may directly influence glycemic response to anti-diabetic interventions, or alternatively be in linkage disequilibrium with other variants with these effects. Further exploration of these genetic factors will increase our understanding of the contributions of glucagon gene products to glucose homeostasis and as modifiers of the effects of metabolic interventions.

857F
Genetic and Environmental Influences on the Age of Onset of Age-Related Macular Degeneration (AMD) in the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA) Cohort. L. Shen1, R. Melles2, S. Sciortino3, D. Ranatunga4, L. Walter5, L. Sakoda6, R. Whitter1, T. Hoffmann7, M. Kvale2, Y. Banda2, N. Risch8, C. Schaefer9, E. Jorgenson9. 1) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 2) Institute for Human Genetics University of California, San Francisco.

Age-Related Macular Degeneration (AMD) is a highly heritable, common disease with a prevalence of 30% in non-Hispanic white adults aged 75 years or older. In addition to genetic factors, environmental risk factors, including smoking behavior, are known to affect the risk of developing AMD. Here we evaluate the effect of previously identified AMD risk variants and environmental risk factors on the age at onset of AMD in the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA) cohort (n=110,266). Cohort members completed a health survey questionnaire that included information on smoking behavior and alcohol consumption. A genetic risk score (GRS) was calculated based on 10 independent risk variants that had been identified in prior studies. During the 8 year follow-up period after the GERA enrollment, 3,278 non-Hispanic white AMD incident cases were identified from the electronic medical record (EMR). These cases had an average of 13.9 years of continuous observation via EMR prior to the first AMD diagnosis, limiting the possibility that subjects might have been diagnosed with AMD at a younger age. The mean onset age was 76.16 years (SD=8.52). Subjects were divided into quintiles based on the distribution of the GRS in 78,697 non-Hispanic white GERA cohort members. Cases in the highest GRS quintile had a mean age of onset 2.44 years younger (p=0.01) than those in the lowest quintile. Current smokers had a 2.34 year earlier onset (p=0.007) than non-smokers. Subjects who exceeded the NIAAA daily safe drinking limit at least once per month at the time of the survey had a 2.14 year (p=7.2*10^-6) earlier age of onset. Stratified analyses revealed an interaction between alcohol consumption and GRS (p=0.0134). Exceeding the NIAAA daily limit was associated with a 3 year earlier age of onset in subjects in the top half of the GRS distribution, whereas age of onset was only 0.8 years earlier in the bottom half. Our findings suggest that these genetic and environmental factors influence the age of onset of AMD, and suggest that it may be possible through lifestyle changes to delay the onset of AMD in subjects at high genetic risk.

858W

Absolute pitch (AP), also known as perfect pitch, is the unusual ability to recognize a pitch without an external reference. The most popular current view is that both environmental (such as an early age at the beginning of musical training) and genetic (as evidenced by familial aggregation and twin studies) factors are involved in the acquisition of the trait. With the goal of disentangling the contribution of genes vs. environment to the manifestation of AP, 89 musicians were subjected to an interview and musical tone identification test. Subjects were recruited using a mixed strategy, which included university music students as well as volunteers who responded to a newspaper article. Therefore, prevalence cannot be estimated. The test consisted of the identification of 40 piano and 40 ‘pure’ tones, with four seconds between tones. Subjects who correctly identified 75 or more tones were considered AP possessors. A total of 16 AP possessors were identified, eight of whom correctly named all 80 tones. None of the AP possessors reported family members with perfect pitch. In most of the cases, however, the parents and siblings of the subjects had no musical training. A new test for AP, which does not rely on the naming of specific tones, could help elucidate whether familial aggregation is present. Interestingly, the average age at the beginning of musical training did not differ between AP (7.8 ± 3.9 years) and non AP (8.8 ± 4.1 years) subjects (t=0.81, d.f.= 87, p=0.42), challenging the assumption that early musical training is a requirement for the development of AP. A total of 56% (9 out of 16) of AP possessors began their musical training after age 6; 6 of these after age 10. A possible interpretation is that in the presence of a strong genetic predisposition, musical training can start at more advanced ages. Additionally, there is an interesting group of 16 non-AP possessors who correctly identified more tones than those predicted by chance and failed to correctly identify tones were off by only a semitone. They represent a middle category between true AP and true non-AP possessors. In agreement with the likely complex nature of perfect pitch, this category possibly includes individuals in whom genetic and environmental exposure was not enough to develop the trait, or the other way around.
859T

Interplay of Genetic Risk (CHRNA5) and Environmental Risk (partner smoking) on Smoking Cessation Success. L. Chen1, T.B. Baker2, M. Munafò1, L.J. Bierut3. 1) Psychiatry, Washington University School of Medicine, St. Louis, MO, USA; 2) Tobacco Research and Intervention, University of Wisconsin, School of Medicine, Madison, WI 53711, USA; 3) Department of Biobehavioral Health & Society, Pennsylvania State University, University Park, PA, USA.

Objective: Identifying predictors affecting individual variation in quitting success is important for both clinical response prediction and public health. Smoking cessation success can be predicted by both genetic factors such as variants in the nicotinic receptor gene (CHRNA5) and environmental factors such as partner smoking. This study tests whether a specific genetic risk for cessation success is moderated by an environmental feature.

Method: In a community-based, longitudinal study (N=1,856) of pregnant women who were smokers before pregnancy, and a randomized comparative effectiveness smoking cessation trial (N=1,065), we examined if the effect of partner smoking on smoking cessation varies with a genetic variant in CHRNA5. Smoking cessation was defined as the trajectory of amount smoked over time in the observational study of pregnant women, and as the trajectory of abstinence among women with complete and available abstinence data in the trial.

Results: While the pregnant women decreased their smoking quantity over time, both the genetic risk (rs16969968(A)) and partner smoking predict heavier smoking during pregnancy. However, rs16969968 interacted with partner smoking: the genetic risk is significantly increased for expectant mothers who have a partner who smokes (b=0.071, 95% CI=0.013-0.13, p=0.017). Similarly, among the smoking cessation trial participants receiving the placebo, a similar interaction was found between rs16969968 and partner smoking: The genetic risk is significantly increased for smokers with a partner who smokes (b=0.20, 95%CI=0.049-0.36, p=0.010). This interaction between genetic and environmental factors occurs only in the placebo group and not in the active pharmacotherapy group (significant interaction, b=0.25, 95% CI=0.42 to 0.091, p=0.0023). Conclusions: The risk of partner smoking and the CHRNA5 genetic marker interact in both studies of smoking cessation. The risk associated with CHRNA5 is moderated by the environment, suggesting that the genetic vulnerability may be mitigated by environmental factors such as partner smoking, with increased risk of cessation failure with combined genetic and environmental risks ameliorated by cessation pharmacotherapy. Incorporating both genetic and environmental factors is critical in designing successful smoking cessation treatments.

860F

Genetic Predisposition for Hypertriglyceridemia is Modified by Extremes of Adiposity. C. Cole1, M. Nikpay2, R. Dent3, R. McPherson4. 1) Atherogenomics Laboratory, Univ Ottawa Heart Institute, Ottawa, ON, Canada; 2) Weight Management Clinic, The Ottawa Hospital, Ottawa, ON, Canada.

We determined the predictive value of a genetic risk score (GRS) for plasma lipids in 2 cohorts of obese and lean subjects genotyped on Affy arrays: 1) SHS School of Public Health, National University of Singapore, Singapore, Singapore; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; 3) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; 4) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore; 5) Department of Medicine, National University of Singapore, Singapore, Singapore; 6) Department of Statistics and Applied Probability, National University of Singapore, Singapore.

861W

Education Influences the Association between Genetic Variants and Refractive Error: A Meta-analysis of Five Singapore Studies. Q. Fan1, R. Wojciechowski2, M.K. Ikram3, C.Y. Cheng3, P. Chen1, X. Zhou1, W. Pan1, C-C. Khor1,4,5, E.S. Tai1, T. Aung1,4,6, T-T. Wong1,4,6, Y-Y. Teo1,4,6, S-M. Saw1,3,1, L.J.Bierut1,2,3, L. Jiang1,3, H. Källberg1, L. Arlésigt1,2, S. Rantapää-Dahlqvist2, L. Kräskoga, L. Padyukova2, L. Alfredsson2, The Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) Study Group. 1) Department of Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Department of Public Health and Clinical Medicine, Rheumatology, Umeå University; 3) Rheumatology Unit, Department of Medicine, Karolinska Institute, Stockholm, Sweden.

Background: Rheumatoid arthritis (RA) is believed to have a multifactorial etiology, involving both genetic and environmental components, and can be divided into two major subsets based on the presence/absence of anti-citrullinated protein/peptide antibodies (ACPAb). Smoking is the best established environmental risk factor. Despite progress from genome-wide association studies (GWAS), identified genetic variants only explain a small proportion of RA occurrence, and gene-environment interaction could add etiologic information. In current study is to investigate large scale gene-environment interaction between smoking and SNPs for each RA subset. Methods: We analyzed data from the Swedish EIRA case-control study using logistic regression models. Smoking history (never vs ever smoking) was collected through questionnaires. Genetic risk was obtained from a custom made Illumina Illumicomp chip. Interaction between smoking and 133648 genetic markers that passed quality control were examined for the two RA subsets (1590 ACPA positive cases, 891 ACPA negative cases; compared with 1856 controls). Attributable proportion (AP) due to interaction was evaluated for each smoking-SNP pair and corrected for multiple comparisons. We performed replication in a case-control study of RA from northern Sweden. To further validate the results we also performed an association analysis using GWAS data for the EIRA study. Results: In ACPA positive RA, 102 SNPs were significantly interacting with smoking after Bonferroni correction, all located in the HLA region (one in HLA class I region, the rest in HLA class II region); 51 were replicated in the independent case-control study from northern Sweden. No additional loci besides from chromosome 6 turned up in the GWAS validation. After adjusting for HLA-DRB1 shared epitope (SE), 15 SNPs remained significant for ACPA positive RA, with 8 of them replicated. For ACPA negative RA, no SNP passed threshold for significance. 10 genes were identified for ACPA positive RA, no SNP passed threshold for significant interactions with education were also observed for axial enthesitis and myopia. Our study shows that low level of education may attenuate the effect of risk alleles on myopia. These findings further underline the role of gene-environment interactions in the pathophysiology of myopia.

Table: Triglycerides

<table>
<thead>
<tr>
<th>Triglycerides</th>
<th>OBLE cohort</th>
<th>adjusted R2</th>
<th>C</th>
<th>adjusted R2</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBLE cohort</td>
<td>beta GRS</td>
<td>mmol/L</td>
<td></td>
<td>beta GRS</td>
<td>mmol/L</td>
</tr>
<tr>
<td>lean</td>
<td>0.0 . 0 9 4</td>
<td>(5.8e-10)</td>
<td>0.042</td>
<td>0.0 . 1 0 2</td>
<td>(4.2e-06)</td>
</tr>
<tr>
<td>obese</td>
<td>0.0 . 1 5 4</td>
<td>(+2e-16)</td>
<td>0.090</td>
<td>0.0 . 1 5 3</td>
<td>(1.5e-07)</td>
</tr>
<tr>
<td>HDL-C</td>
<td>beta GRS</td>
<td>mmol/L</td>
<td></td>
<td>beta GRS</td>
<td>mmol/L</td>
</tr>
<tr>
<td>lean</td>
<td>0.0 . 5 8 9</td>
<td>(+2e-16)</td>
<td>0.083</td>
<td>0.0 . 6 8 4</td>
<td>(1.0e-08)</td>
</tr>
<tr>
<td>obese</td>
<td>0.0 . 4 9 5</td>
<td>(4.3e-14)</td>
<td>0.060</td>
<td>0.0 . 5 3 5</td>
<td>(6.6e-05)</td>
</tr>
</tbody>
</table>
863F

A study on genetic variation associated with visceral adipose tissue and interaction of life style on the expression of genetic variation. H. Kwon1,2, K. Sohn1, B. Cho1, H. Choi1. 1) Seoul National University Hospital, Seoul, South Korea; 2) Healthcare Research Institute, Seoul, South Korea.

Introduction: Obesity, especially abdominal obesity is known as a risk factor for various diseases, and also to increase mortality. Among components of abdominal obesity, visceral adipose tissue (VAT) is a risk for various metabolic diseases and mortality. Genetic variations affecting abdominal obesity have been identified, and interactions between genetic variations and lifestyle factors also have been studied. But there has been no study on genetic variations altering VAT in Koreans, nor study on interaction between genetic variations and lifestyle factors affecting VAT in Koreans. In this study, we have tried to find the association of VAT and SNPs which were previously known to be associated with obesity, and also to test the interaction of life style such as smoking, alcohol drinking and physical activity on the association of VAT and SNPs. Methods: I have selected 17 SNPs previously known to be associated with obesity, and analyzed selected SNPs through realtime PCR from the blood of Korean men aged 20 to 65 who took comprehensive health checkup programs including abdominal fat analysis with CT scan. I have tested the association between selected SNPs and obesity traits including VAT with linear regression analysis, and also checked interaction of life style factors on the association. I also tried stratified analysis with the criteria of body mass index 25kg/m2. Results: Rs9939609 was associated with VAT in dominant model, with AT/AA allele type tends to 6.8cm2 higher in VAT, 0.6kg/m2 higher in body mass index, 1.4cm thicker in waist circumference, 0.7% more in body fat, and 11.2cm2 higher in abdominal subcutaneous fat. On interaction analysis, rs1514175, rs2112347, rs1077664 were affected by alcohol drinking, and rs1718537 and rs3817334 were affected by physical activity on their association with VAT, but on stratified analysis by life style factors there was no significant difference in VAT according to allele type. On stratified analysis by body mass index, rs713586 and rs3810291 were associated with VAT only in subjects with body mass index higher than 24kg/m2 Conclusion: Rs9939609, rs713586 and rs3810291 were associated with VAT in Korean male, and this finding could be implemented in practice and study in the field of obesity.

864W

Lineages based genome-wide association analysis in tuberculosis. S. Mahasirimongkol1, N. Smitipat1, T. Juthayothin2, T. Mushiroda3, S. Wattanapokayak1, N. Wichukchinda1, S. Neduswan4, K. Dokladda2, K. Rukserere5, P. Billamas6, P. Palittapongarnpim5, B. Chaiyasiriroj6, A. Chaiprasert7, H. Yana8,9, K. Tokunaga2. 1) Medical genetics section, National Institute of Health, Nonthaburi, Thailand; 2) National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathumtani; 3) Laboratory for Pharmacogenetics, Center for Integrative Medical Sciences, RIKEN, Japan; 4) Chiangrai prachanukroh hospital, Chiangrai, Thailand; 5) Department of Microbiology, Faculty of Science, Mahidol University; 6) TB/HIV research foundation, Chiangrai, Thailand; 7) Department of Microbiology, Faculty of Medicine Siriraj Hospital, Bangkok, Thailand; 8) Division of Medical Diagnostics, Department of Clinical Laboratory, Fukujuki Hospital, Japan; 9) Department of Human Genetics, Graduate School of International Health, Tokyo, Japan.

Objectives: Genome wide association of tuberculosis in Asians identified a locus near MAFB associated with young tuberculosis (Mahasirimongkol, Yanai et al. 2012). Beijing or Modern lineages of M.tuberculosis is prevalent in young tuberculosis in Thailand, and genetic heterogeneity based on pathogen diversity in M. tuberculosis is a potential confounding factor contributing to inconsistent association evidences in genetic epidemiology of tuberculosis across populations. Global spatial distribution of M. tuberculosis lineages support plausible co-evolution with human populations(Gagneux 2012). Two major lineages of Beijing (belong to Modern) and East African Indians (EAI) belong to Ancient circulating in the Thai tuberculosis populations, allowing study of lineages based association in tuberculosis; we did genome wide association analysis based on two major M.tuberculosis lineages to determine the lineage specific association in tuberculosis. Methods: M. tuberculosis were genotyped using the large sequence polymorphisms (LSP)-based PCR and direct repeats (DR)-based spoligotyping methods. The genome-wide genotyping was carried out with the Illumina 610 array following manufacturer protocol, data analysis was carried out with GenABEL. List of 242 TB candidate genes were retrieved from the HugeNet database. The association analysis specific to modern strains and ancient strains were carried out and reported.

Results: In this analysis, 201 Modern TB and 204 Ancient TB were available for the genome-wide association analysis. The comparisons were made against each group of TB, in each comparison; the additional 835 controls genetically matched samples were used as control group. For this analysis, QQ plots, Manhattan plots and regional association plots were analyzed and presented.

Conclusion: Candidate genes analysis supported genetic heterogeneity based on lineages of TB. The most interesting association from this analysis is the heterogeneity within MHQ class II in TB. Among the top 50 SNPs, a locus associated with Ancient TB near Leukocyte receptor cluster (LRC) was identified, replication analysis of these findings are required for concrete confirmatory evidences.
FTO variants, dietary intake, and body mass index: results from 865T

FTO genotype, dietary intake, and body mass index: results from 865T

FTO is the strongest known genetic susceptibility locus for obesity. Experimental studies in animals suggest the potential roles of FTO in regulating food intake and energy expenditure. The interactive relation among FTO variants, dietary intake of energy-dense macronutrients, and body mass index (BMI) are complex and results from previous often small-scale studies in humans are highly inconsistent. We performed large-scale analyses based on cross-sectional data from 177,330 adults from 40 studies according to a standardized analytical plan to examine: 1) the association between the FTO-9939609 variant (or a proxy SNP) and total energy and macronutrient intake; and 2) the interaction between FTO variant and dietary intake on BMI. Macronutrient intake was expressed as the percentage of total energy intake. Dietary variables were dichotomized into two categorical variables based on the median intake in each study for the interaction analysis. The BMI-increasing allele of the FTO variant showed a significant association with higher dietary protein intake (effect per allele =0.08, 95% CI 0.06, 0.10), lower total energy intake [-6.4 [-10.1, -2.8] kcal/day], P=0.001), and lower dietary carbohydrate intake [-0.07 [-0.11, -0.02]%, P=0.004]. The association between FTO variant and protein intake (P =7.5x10-9) remained significant after adjustment for BMI. We did not find significant interactions between the FTO variant and dietary intake of total energy (P=0.25), protein (P=0.87), carbohydrate (P=0.80), or fat (P=0.13). The FTO variant was similar between participants in the low intake and the high intake groups of these dietary factors. In conclusion, the BMI-increasing allele of FTO seems to be associated with increased dietary protein intake. There is no obvious evidence supporting interactions between the FTO genetic variant and intake of energy or macronutrients in relation to BMI.

Genome-wide environmental interaction (GWEI) analysis using multi-dimensional data reduction principles to identify asthma pharmacogenetic loci in relation to corticosteroid therapy. F. Van Lishout1,2, QL. Duan3, K. Tantishira3,4, K. Van Steen3,4.

Genome-wide environmental interaction (GxE) and gene-gene (GxG) interaction studies share a lot of challenges via the common genetic component they involve. GWEI studies may therefore benefit from the abundance of methodologies that are available in the context of genome-wide epistasis detection methods. One of these is Model-Based Multifactor Dimensionality Reduction (MB-MDR), which does not make any assumption about the genetic inheritance model. MB-MDR involves reducing a high-dimensional GxE space to a GxE factor levels that either exhibit high or low or no evidence for their association to disease outcome. In contrast to logistic regression and random forests, MB-MDR can be used to detect GxE interactions in the absence of any main effects or when sample sizes are too small to be able to model all main and GxE interaction effects. In this ongoing study we demonstrate the opportunities and challenges of MB-MDR for genome-wide GxE interaction analysis and analyzed the difference in prebronchodilator FEV1 following 8 weeks of inhaled corticosteroid therapy, for 565 Caucasians from the Childhood Asthma Management Program (CAMP). In particular, we first followed standard marker quality control procedures. Missing genotypes were imputed with MaCH using 1000 Genomes Project Reference Panels (8,521,072 SNPs). Second, residuals were computed based on the polygenic regression model to correct for different types of relatedness and population stratification and to account for possible confounders such as age and height at baseline and gender. Third, LD pruning with maximum correlation threshold 0.5 between markers of 0.5 was applied yielding 867,859 SNPs. Third, we curtailed MB-MDR to perform a GxE interaction analysis, with E dichotomous variable coding for corticosteroid therapy (1: inhaled corticosteroids; 0: other). Amongst our top results, we found 3 SNPs mapped to ANGPT2 and MCHP1, related to the vascular remodeling and DNA damage repair. c) rs108988318 mapped to SLC22A1, a member of the solute carrier organic anion transporter family identified as participant of neurodifferentiation. b) rs2515471 mapped to ANGPT2 and MCHP1, related to the vascular remodeling and DNA damage repair. c) rs10898518 mapped to SLC22A1, a member of the solute carrier organic anion transporter family involved in signal propagation in synapses and cell signaling. Follow-up of these results include replication in independent datasets and formal testing and validation of identified interactions as corticosteroid-related pharmacogenetic loci.
Influence of physical activity on body mass index in relation to well-replicated obesity loci in African-American adults: The ARIC Study.

K.L. Young, A.E. Justice, E.W. Demerath, W.-L. Kao, E. Boerwinkle, K.E. North, K.L. Monda. 1 Carolina Population Center, University of North Carolina, Chapel Hill, NC; 2 Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 3 Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC; 4 Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN; 5 Department of Epidemiology, Johns Hopkins University, Baltimore, MD; 6 Division of Epidemiology, School of Public Health, University of Texas Health Science Center, Houston, TX; 7 Center for Observational Research, Amgen, Inc, Thousand Oaks, CA.

Obesity is a known risk factor for many chronic diseases, and physical activity (PA) is often used to lower those risks by reducing weight. However, the impact of PA on obesity exhibits considerable variation, which is likely influenced by genetics. African Americans (AA) are at particularly high risk of obesity, but remain understudied. We investigate the modifying influence of PA on 24 BMI loci which generalize or for which a better SNP in the same locus has been reported in AA using the Atherosclerosis Risk in Communities (ARIC) Study, a population-based sample of adults aged 45-64 years at baseline. Our analyses were restricted to genotyped individuals of self-reported African descent (N=2,373). BMI (kg/m2) was computed from measured height and weight. PA was measured using a modified Baecke questionnaire, resulting in an ordinal score (range: 1-10). We assumed an additive genetic model, regressing BMI on SNP, PA level, SNP by PA, sex, age, current smoking, education, center, alcohol consumption, and principal components (PCs) to account for ancestry. Additionally, we performed main effects analyses controlling for sex, age, current smoking, center, and PCs. The majority of our sample was female (63%), with a mean age of 53.3 (SD 5.7) years, and at least a high school education. Mean BMI was 29.7 (SD 6.0) kg/m2, and PA was 6.98 units (SD 1.19, range: 2.0-9.0). Ten of 24 SNPs were those with low PA increase with additional copies of the G (risk) allele of 0.12 kg/m2 for high PA individuals (p-interaction=0.045). Mean BMIs for low PA individuals resulted in an increase of 0.04 kg/m2 for low PA individuals, but a decrease of 0.02 kg/m2 for high PA individuals (p<0.05). Each additional copy of the risk allele in MAP2K5 reached suggestive significance (p<0.002), MAP2K5 reached suggestive significance (p<0.002), with at least one known viral respiratory pathogen e.g., human rhinovirus (HRV) or respiratory syncytial virus (RSV). Gene expression profiles showed highly similar patterns among the major subgroups. There was a dramatic up-regulation of interferon pathway and innate immunity on the first day of infection. This usually persisted for 2 days. A convalescent phase was observed on days 4 and 6 after infection. By day 21 the gene expression pattern had returned to levels indistinguishable from baseline. Using lineage and activation state specific transcripts to produce cell decomposition scores, patterns of acute depression of the B and T lymphocytes were observed accompanied by the evidence of dramatic activation of dendritic cell (DC) and T cells. Transcriptional profiling gives a genome wide view of a coordinated systemic response to acute viral respiratory infection. There are two clear phases of gene expression, corresponding to intense activation of innate immunity pathways followed by a convalescent phase marked by cell proliferation and repair.

Transcriptional response to acute viral respiratory infection - a prospective cohort study. Y. Zhai, J. Belmont, R. Atmar, J. Quailes, N. Arden, K. Bucassas, J. Wells, D. Niño, X. Wang, G. Zapata, C. Shaw, L. Franco, R. Couch. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Medicine, Baylor College of Medicine, Houston, TX; 3) Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX; 4) Children’s Nutrition Research Center, Baylor College of Medicine, Houston, TX; 5) Department of Microbial and Molecular Pathogenesis, Texas A&M University System Health Science Center, College Station, TX; 6) Department of Pediatrics, Baylor College of Medicine, Houston, TX.

Objectives: Acute viral respiratory infections (ARIs) are responsible for a large number of outpatient visits and hospitalizations in the U.S., and the threat of pandemic influenza will likely add to these numbers. A new influenza A(H1N1) virus, which emerged in April 2009, spread worldwide and continued to circulate in the following year. In this study we aim to identify functional gene networks and pathways that are correlated with immune response to influenza infection by time-course analysis of the host transcriptional profiles. Methods: 1618 healthy adults, enrolled in fall 2009 and 2010, were followed for acute viral respiratory disease. Subjects reporting moderate to severe acute respiratory illness had virus quantitation for 3 weeks. Peripheral blood samples for RNA extraction were obtained on the first day of illness (day 0) and then on 2, 4, 6, and 21 days after first symptoms. In addition to an RNA sample obtained at enrollment, a final RNA sample was obtained at the end of the study. RNA samples were analyzed using expression microarrays and the patterns of gene expression were analyzed to identify differentially expressed transcripts. Result: Among the 133 subsets of the ARI cohort, subjects who had viral respiratory infection completed 21 day study visits. Our results suggest that unique patterns are observed concomitant with viral infection. The expression profiles were consistent across the major subgroups, consistent with the expectation of the subgroups.

Insights into the molecular arms race between the Malaria parasite and the human host from genomic analysis of over 15,000 African individuals. C.C.A. Spencer, L.S. Quang, G. Bandt, K. Rockett, D. Kwiatkowski, MalariaGEN. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

The continuing burden of death and illness that malaria inflicts in endemic regions underlines its likely important role in human evolution. To gain insights into the key molecular determinants of human susceptibility to severe malaria we studied over 15,000 individuals from eight countries across sub-Saharan Africa, using genome-wide SNP genotyping, as part of the MalariaGEN consortium (www.malariaigen.net). We use these data, in combination with novel statistical methodology, to jointly describe patterns of association and natural selection across the genome. By systematically coupling signals of association with models of recent evolution across genes of different function we can test specific hypotheses. We assess the evidence for two broad classes of protective effects: the first, protective alleles that are maintained under balancing selection through deleterious effect (for example sickle cell disease) which have consistent effect populations; the second, protective alleles that are under frequency-dependent selection, and are expected to have different effects on susceptibility across populations, due to an interaction with parasite diversity. These new observations reinforce the role of red blood cell surface proteins as important in severe malaria susceptibility, putatively by inhibiting parasite invasion mechanisms. In addition, our genome-wide data can be used to assess the relationship between recently reported shared human and chimpanzee polymorphisms and the malarial parasite disease caused by the malaria parasite. As well as helping to quantify the impact of malaria susceptibility alleles on human diversity in Africa, we are also able to use signals of natural selection to aid identification of new susceptibility loci.
The MTHFD1 1958 G>A (R653Q) variant is associated with elevated C-reactive protein and body mass index in a Canadian premature birth cohort. K.E. Christensen1,2, J.J. Roedig1,2, M. Dahhou3, M.S. Kramer1, R. Rozon1,2, 1) Dept. of Human Genetics, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 2) Dept. of Pediatrics, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 3) Dept. of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada.

MTHFD1 1958 G>A (rs2236225) is a non-synonymous variant in the synthetase domain of the trifunctional MTHFD1 enzyme. It reduces stability and impairs the novel purine synthesis (factor for C-reactive protein, CRP) as a model for this variant. The majority of the mice were healthy and normal; however, the rate of developmental defects was higher in offspring of Mthfd1S+/− dams. In addition, nonpregnant female Mthfd1S+/− mice were heavier than their wild-type littersmates, and while white blood cell counts (specifically neutrophils) were reduced, particularly during pregnancy. To examine immune function and weight in women with the 1958 G>A SNP, we genotyped 206 women with spontaneous preterm births and 443 controls from a large multicentre cohort of Quebec women. There was no association with prematurity, with or without stratification by median plasma folate (29.6 nM). In grouped cases and controls, plasma C-reactive protein (CRP, an immune marker associated with low-grade inflammation) was increased due to the A allele (p=0.053, ANOVA; p for trend = 0.041) in women with folate intake is low. Additional cohorts should be evaluated, particularly in the Irish population, which is not folate fortified. We have previously found that: 1) Carriers of the A allele have a lower BMI, and 2) the rate of developmental defects was higher in offspring of high versus low CRP including genotype, maternal age, smoking, BMI and education in the model, CRP was increased in GG compared to AA women with folate intake is low. (p=0.055). There was also a strong association between pre-pregnancy BMI >30 and CRP in this analysis, both above and below median folate (p<0.001). A nonsignificant dose-response relationship was observed between mean BMI and genotype, especially in the low-folate group, when analyzed by ANOVA. However, there was a significant association between the 1958G>A variant and BMI, when examined for low (<18.6) and high (>24.9) BMI (p<0.03). The inheritance of obstructive sleep apnea (OSA) is complex. Previously, APOE-e4 and increased facial convexity (mandibular retrusion, i.e., a Class II (CII) sagittal skeletal jaw relationship) have independently been associated with OSA. We hypothesized that in OSA there is an association between APOE-e4 and non-CII compared to CII subjects. Associations between APOE-e4 with BMI, and Apnea-Hypopnea-Index (AHI); and jaw relationship with BMI and AHI were also studied. Materials and Methods: Seventy-six Caucasian OSA subjects with an AHI > 15 were classified into different sagittal skeletal jaw relationships by oral exam and profile photos. In addition to CII, Non-CII includes Class I with straight profile, and Class III with mandibular protrusion. DNA via saliva was extracted. Two SNPs (rs429358 and rs7412) were genotyped to determine if any associations with outcomes of interest. A chi-square analysis assessed Hardy-Weinberg-Equilibrium (significance at p<0.05). Fisher’s exact test was used to look for association between APOE-e4 status and the skeletal classification of OSA subjects. ANOVA compared BMI and AHI among the skeletal types. Results: Twenty-eight CII subjects (mean BMI 30.7, AHI 33.6) and fifty-one non-CII subjects (mean BMI 37.4, AHI 44.2) were examined. Seventy subjects were obese, and six subjects had a normal BMI. All six individuals with a normal BMI were CII. APOE-e4 was not associated with different sagittal facial profiles. CII subjects had significantly lower BMI (CII 37.2 vs CII 37.7, p<0.05) and BMI close to normal; however, the rate of developmental defects was higher in offspring of OSA with or w/out increased BMI. No association between non CII and APOE-e4 was found in this preliminary study.
Occupational exposures to potential irritants and asthma: effect modification by glutathione S-transferase Z1 and ATP-binding cassette transporters polymorphisms. M. RAVAT, I. AHMED, O. DAMAS, M. KÖGEVINAS, N. PROBST-HENSCHEL, P. TUBERT-BITTNER, N. LÉ MOUAL, F. DEMENAIAS, R. NADIC, A. OCHS. 1) INSERM U1018, Centre for research in Epidemiology and Population Health (CESP), Respiratory and Environmental Epidemiology Team, Villejuif, France; 2) Univ Paris-Sud 11, UMRS 1018; 3) INSERM U1018, Centre for research in Epidemiology and Population Health (CESP), Biostatistics Team, F-94807, Villejuif, France; 4) Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; 5) Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute Swiss TPH, Switzerland; 6) University of Basel, Switzerland; 7) INSERM, UMR-S946, F-75010, Paris, France; 8) Univ Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d’Hématologie, F-75007, Paris, France; 9) Fondation Jean Dausset-Centre d’Etude du Polymorphisme Humain (CEPH), Paris, F-75010, France.

Occupational asthma (OA) is a good model to study the pathophysiology of asthma. Around 15% of adult asthma would be caused by occupational exposures, and more than 400 distinct agents have been identified as causing OA. The mechanisms of asthma induced by potential irritants, such as cleaning and chemical products, are unclear but could be related to oxidative/nitrosative stress. Our goal was to identify polymorphisms (SNPs) in genes involved in the response to oxidative/nitrosative stress interacting with occupational exposures known to have an effect on asthma, in the familial case-control French Epidemiological study of the Genetic and Environmental factors of Asthma (EGA, https://ega.net embassy.insERM.fr/). A large set of genes (162 genes, 4979 SNPs) was selected according to a pathway-based strategy that integrates biological knowledge related to occupational exposures. Occupational exposure to potential irritants was evaluated using the asthma job exposure matrix (Asthma JEM, http://ceesp.cefay.embassy.insERM.fr/asthhamjer) and job-specific questionnaires for cleaners and healthcare workers. First, the marginal genetic associations and the job-specific associations were estimated using a GEE logistic regression models, adjusted for age, sex and principal components of population ancestry. Then, the marginal genetic associations and the occupational exposures interactions were estimated separately using GEE logistic regression models, adjusted for age, sex and principal components of population ancestry. The mechanisms of asthma induced by potential irritants, such as cleaning and chemical products, are unclear but could be related to oxidative/nitrosative stress. Our goal was to identify polymorphisms (SNPs) in genes involved in the response to oxidative/nitrosative stress interacting with occupational exposures known to have an effect on asthma, in the familial case-control French Epidemiological study of the Genetic and Environmental factors of Asthma (EGA, https://ega.net embassy.insERM.fr/). A large set of genes (162 genes, 4979 SNPs) was selected according to a pathway-based strategy that integrates biological knowledge related to occupational exposures. Occupational exposure to potential irritants was evaluated using the asthma job exposure matrix (Asthma JEM, http://ceesp.cefay.embassy.insERM.fr/asthhamjer) and job-specific questionnaires for cleaners and healthcare workers. First, the marginal genetic associations and the job-specific associations were estimated using a GEE logistic regression models, adjusted for age, sex and principal components of population ancestry. Then, the marginal genetic associations and the occupational exposures interactions were estimated separately using GEE logistic regression models, adjusted for age, sex and principal components of population ancestry.
878F
The IRF5-TNPO3 association has two components in systemic lupus erythematosus (SLE), which are shared with other autoimmune disorders. Recent improvements in genetic methods allow for more complete identification of candidate causal variants. Exploiting genotyping, DNA sequencing, imputation, and trans-ancestral mapping, we modeled the IRF5-TNPO3 genetic association on chromosome 7, now implicated in two immunotherapies and autoimmune disease susceptibility. We identified genetic variants across five ethnicities in 8,395 systemic lupus erythematosus (SLE) cases and 7,367 controls. We resolve separate lupus-risk associations in the IRF5 promoter (all ancestries) and an extended European haplotype-strongly associated with lupus in one European decent but non-polymorphic in subjects from Asia or Sub-Saharan Africa. Adjusting for two variants tagging each genetic effect removes the association of other variants in the region. Using both frequentist and Bayesian approaches, we identify the mostly likely causative variants to three IRF5 promoter variants (confined to 5.7 kb) and 22 variants of the 85.5 kb European haplotype that spans IRF5 and TNPO3. The possible statistical models from this sample virtually eliminate the previous purported IRF5 functional variants as causal. Strikingly, this model also appears to operate in Sjögren’s syndrome and systemic sclerosis (both components) and primary biliary cirrhosis (haplotype only), demonstrating the nuances of similarity and difference in autoimmune disease risk mechanisms at IRF5-TNPO3. Finally, we performed case-only subphenotypic analyses to identify variants in the IRF5-TNPO3 region uniquely associated with particular phenotypes of SLE. We found strong association with age-of-onset (variants in the haplotype) and the antigen specificity of autoantibodies anti-Ro and anti-.dsDNA (variants in the IRF5 promoter). In conclusion, genetic variants in the IRF5-TNPO3 region can discriminate the diverse clinical presentations of lupus and the clinical presentation of lupus. Given the importance of this locus to multiple immune disorders, future studies will be aimed at understanding the mechanisms driving the genetic associations. (72 other collaborators, providers of samples, genetic, analytical, and financial resources will be provided at presentation.).

879W
Epiregulin (EREG) and human V-ATPase (TCIRG1) are not associated with pulmonary tuberculosis in West Africans. J.B. Harley1,2, J.C. Bissau1, The Gambia. S.M. Williams1,2, M.J. White1,2, A. Tacconelli1, C. Wejsa3, P.C. Hill3, G. Novelli3, G. Sirugo3, 1) Department of Genetics, Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Centro di Genetica, Ospedale San Pietro FBF, Rome, Italy; 4) Bandim Health Project, Danish Epidemiology Science Centre and Statens Serum Institute, Bissau; 5) Centre for International Health, University of Oslo School of Medicine, Oslo, Norway. Tuberculosis (TB) is caused by Mycobacterium tuberculosis (M. tuberculosis) and infects approximately one third of the world’s population, with the majority of infection occurring in Africa and South-east Asia. Active TB, of which pulmonary TB (PTB) accounts for approximately 90%, has one of the highest mortality rates for an infectious disease. There is a large discrepancy between latent TB infection and incidence of active TB, with only a small proportion of latent TB infection progressing to active TB. Recent studies have identified two novel TB susceptibility candidate genes, EREG (Epiregulin) and TCIRG1 (human ATPase, a3 isofrom), in Vietnamese and Italian populations, respectively. The main aim of this study was to validate the associations between EREG and TCIRG1 with PTB in at risk populations from West African endemic areas. Using a discovery cohort from Guinea Bissau (n = 289 cases, 322 controls) and a replication cohort from the Gambia (n = 240 cases, 248 controls), we assessed association with 14 SNPs in EREG and TCIRG1; using single and multi-locus analysis to determine whether these variants generalized as a susceptibility factor for PTB. Four variants in EREG and two variants in TCIRG1 were identified as susceptibility variants. Logistic regression analysis revealed one SNP in TCIRG1, rs10896289, as significantly associated with PTB (p = 0.042) and another SNP in EREG, rs1563826, that was marginally associated with PTB (p = 0.056) in the discovery cohort: these associations were not observed in the Gambian replication cohort. Haplotype analyses were also performed to investigate associations with PTB; no significant results were found in either cohort. We also performed multi-factor dimensionality (MDR) analysis to elucidate interaction effects, between the 14 SNPs and other factors. Interestingly, although MDR identified the same variables in both our discovery and replication cohorts, further investigation revealed that the risk models differed. In conclusion, despite being adequately powered to detect the effect sizes previously reported, our single and multi-locus analyses models differed. In conclusion, despite being adequately powered to detect the effect sizes previously reported, our single and multi-locus analyses revealed one SNP in TCIRG1 with genetic variation in these genes may be influenced by inter-population differences in genetic or environmental context and/or the mycobacterial lineage that cause active disease.

880T
Exploring genetic load of known multiple sclerosis risk alleles in Hispanic whites. J. McCauley1, A. Hadjixenofontos1, C.P. Manrique1, A.H. Beecham1, L. Konidari2, P.L. Whitehead3, P.A. Gourraud4, M.A. Pericak-Vance1, L. Torres5, M. Ortega5, K.W. Ramn ahans5, S.R. Delgado5, 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Neurology, School of Medicine, University of California at San Francisco, San Francisco, CA, USA; 3) Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA.

Multiple Sclerosis (MS) is a demyelinating disease with autoimmune etiology and variable expression across populations. Studies on the genetic susceptibility to MS have primarily focused on Northern European (NE) populations, largely due to the increased prevalence in these countries. However, Hispanic patients are underrepresented in well-powered studies of disease presentation, and even more so in investigations into the underlying genetic architecture. Our previous studies have revealed potentially important clinical differences between Hispanic white and non-Hispanic white patients that warrant further investigation. As a first step we have calculated an MS risk score to capture the genetic load attributed to variants identified in NE populations, therefore we hypothesize that an admixed population will be characterized by a partly-overlapping set of genetic variants. The most recent study by the International MS Genetics Consortium has significantly increased the list of known MS risk variants while also fine-mapping previously discovered signals. We use 113 of these variants to construct a weighted MS genetic risk score. Our goal is to assess the extent to which the NE derived MS risk score is able to distinguish Hispanic white cases (n = 186) from non-Hispanic white cases (n = 161). Preliminary results suggest the disease progression due to the variant included in the MS risk score is similar in Hispanic white cases (mean = 12.94, SD = 0.75) and non-Hispanic white cases (mean = 12.91, SD = 0.79) (t-test p-value = 0.67). A more comprehensive examination of the variation present in the known MS risk loci in Hispanic populations is underway. Our study highlights the need for further studies of MS risk factors among Hispanics with the goal of understanding the causes of the differences we see in MS presentation in this population.

881F
Genetic Basis of Height and Skin Pigmentation in Southern Africa. B.M. Henn1,2, J.M. Granka2, A.R. Martin3, C.R. Gignoux4, M. Lin5, J.M. Kidd6, E.G. Hoal7, M.W. Feldman8, C.D. Bustamante9, 1) Dept of Ecology and Evolution, Stony Brook University, Stony Brook, NY; 2) Dept. of Genetics, Stanford University, Stanford CA; 3) Dept. of Biology, Stanford University, Stanford, CA; 4) University California, San Francisco, CA; 5) Dept. of Human Genetics, University of Michigan, Ann Arbor, MI; 6) Molecular Biology and Human Genetics, Stellenbosch University, Tygerberg, South Africa.

No two traits are as immediately recognizable for their variability among and within human populations as height and skin pigmentation. Many evolutionary hypotheses have been proposed to explain their variability, but relatively little is known about the genetic basis of these traits across human populations. In particular, the genetic basis of these phenotypes in the Khoesan populations of southern Africa remain completely unknown. For the first time, we analyze the heritability of height and skin pigmentation in over 200 Xhomas Bushmen and Nama individuals from South Africa, and perform genome-wide association analyses for these traits with both high-coverage exome and SNP array data. We contrast estimates of heritability for height and skin pigmentation and find that most variation in innate skin pigmentation is potentially explained by common single nucleotide polymorphisms (SNPs); height is only moderately heritable and tanning is largely environmental. While varying amounts of European, West African and Khoesan ancestry strongly correlate with skin pigmentation, ancestry does not appear to contribute significantly to variation in height in our samples. Finally, after controlling for kinship relationships in our data, we identify several large effect loci associated with height located near genes previously identified to be associated with body mass index/height (e.g. SLC1A3, VWA8, DGKH). Our results emphasize that analyses of phenotypically and genetically distinct populations in genetic or environmental context may provide new insights into the genetic basis and evolutionary history of complex traits.
882W

Explicit modeling of genetic ancestry improves polygenic prediction accuracy. C. Chen1, J. Han2,3, D. Hunter1,4, R. Kraft1,5, M.刻ma4, A. Price1,2,5

1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Department of Dermatology, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA; 4) Channing Division of Network Medicine, Brigham and Women’s Hospital, Boston, MA; 5) Broad Institute of Harvard and MIT, Cambridge, MA.

Polygenic prediction using genome-wide SNPs can provide higher prediction accuracy than prediction using only known associated SNPs. It is widely known that the polygenic architecture of polygenic disease populations may be partly due to genetic ancestry. However, we hypothesized that explicitly modeling ancestry could improve polygenic prediction accuracy, by providing unbiased estimates for the polygenic component while incorporating associations between ancestry and phenotype as a separate component in the prediction model. This prevents ancestry effects from entering into each SNP effect and being over-weighted. We analyzed three large GWAS of hair color, tanning ability and basal cell carcinoma (BCC) in European-Americans (sample sizes: 7,400 to 8,900) to demonstrate our approach, restricting to a set of 70,000 independent SNPs. We compared polygenic prediction without correction for ancestry to polygenic prediction with ancestry as a separate component in the model, using the top 5 principal components to model ancestry. In 10-fold cross-validation, the $R^2$ for hair color increased by 66% (from 0.0456 to 0.0755; $p$-value $< 10^{-12}$) when explicitly modeling ancestry in the polygenic prediction model. Similarly, the $R^2$ for tanning ability increased by 123% (from 0.0154 to 0.0344; $p$-value $< 10^{-8}$) and the liability-scale $R^2$ for BCC increased by 66% (from 0.0138 to 0.0232; $p$-value $< 10^{-12}$) when explicitly modeling ancestry in a polygenic prediction model. For each of the three traits, an improvement in prediction accuracy remains when including known associated SNPs in the polygenic model. In summary, our results show that explicitly modeling ancestry can be important in polygenic prediction.

883T

Powerful detection of osteoarthritis susceptibility loci by comprehensive examination of clinically important endophenotypes. K. Panoutsopoulou1,2, S. Thanigrachai1, A.G. Day-Williams1,2, L. Southam1,2, K. Hatzikotoulas2, A. Matchan1, M. Doherty1,4, J.M. Wilkinson2, E. Zeggini1, arcOGEN Consortium. 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Academic Unit of Bone Metabolism, Department of Human Metabolism, University of Sheffield, Sheffield, United Kingdom; 3) Biogen Idec, Cambridge, MA, USA; 4) Academic Rheumatology, Nottingham City Hospital, Nottingham, UK.

Osteoarthritis (OA) is a highly heterogeneous disease characterised by variable clinical features with possibly different genetic aetologies. Thus far, the few genetic variants that have been robustly associated with OA ($n=13$ in Europeans) explain only a small proportion of its heritability. Studies performed to date have used very broad phenotypic definitions of OA which do not take into account the differences in pattern of joint involvement between individuals that may represent different physiological processes, and may thus have different genetic aetologies. In a disease like OA the use of expanded and tighter phenotype definitions closer to the biology of the disease may improve objective phenotyping and help to identify loci with stronger associations. We have been part of the largest genome-wide association study for OA (arcOGEN GWAS) and have previously established 9 OA associated loci using broad definitions of OA. Here we have comprehensively analysed an expanded set of narrower, clinically relevant OA endophenotypes derived from radiographs of 2,000 knee and 2,000 hip OA cases from arcOGEN. Variables studied relate to joint morphology, specific anatomic pattern of joint involvement, severity and bone response. Following 1000 Genomes Project-based imputation and stringent quality control, 77 million variants were tested for association with each phenotype. Our results indicate that the study of endophenotypes in OA has the potential to dramatically enhance power to detect OA-relevant associations. For example analysis by knee compartment involvement vs population-based controls yielded 25 independent loci for knee OA at $p<1\times10^{-5}$ vs 1 locus detected for knee OA vs controls in the equivalent binary trait GWAS. In hip OA endophenotype analyses several promising signals were identified some of which are found near genes that are very plausible biological candidates for OA: in the analysis of knee OA with trochanteric and acetabular region involvement GPR98 (OR [95% CI]: $2.03[1.57-2.63], p=2.5\times10^{-8}$) was detected in GPR98. pgr98 knockout mice have a low bone mass phenotype, and a GPR98 polymorphism has recently been associated with osteoporotic fracture in Japanese women. Pattern of hip migration shows strong association with variants in LRCH1 ($p=2.9\times10^{-7}$) previously suggestive associated with OA and BMP1 ($p=2.9\times10^{-7}$) which induces bone and cartilage development. This work has shown an unexpected potential of information on OA susceptibility genes.

884F

Two novel susceptibility loci to Takayasu arteritis and synergistic role of the IL12B and HLA-B regions in a Japanese population. C. TERAO1,2, H. YOSHIZAWA2, A. KISTLER1, T. MATSUURA1, K. CORUMRA1, M. TAKASHI1, M. SHIMIZU1, T. KAWAGUCHI1, Z. CHEN1, T. NARUSE3, A. SATO-OTSURO5, Y. EBANA6, Y. MAEJIMA7, Y. WADA8, J. MARITA8, Y. KAWAGUCHI9, H. YAMANAKA1, S. OGAWA1, I. KOMURO9, R. NAGAI10, R. YAMAO10, Y. TABARA10, M. ISOBE10, T. MIMORI10, F. MATSUDA10, Kyoto University Takayasu Arteritis Consortium. 1) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan; 3) Department of Genetics, Graduate School of Medicine, Kyoto, Japan; 4) Department of Cardiovascular Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 5) Cancer Genomics Project, Graduate School of Medicine, Kyoto University; 6) Department of Bio-informatical Pharmacology, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 7) Department of Cardiovascular Medicine, Tokyo Medical and Dental University, Tokyo, Japan; 8) Division of Clinical Nephrology and Rheumatology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan; 9) Institute of Rheumatology, Tokyo Women’s Medical University, Tokyo, Japan; 10) Jichi Medical University, Tochigi, Japan.

Takayasu arteritis (TAK) is an autoimmune systemic vasculitis of which etiology is unknown. While previous studies have revealed that HLA-B*52:01 has an effect on TAK susceptibility, no other genetic determinants have been established so far. Here, we performed genome-scanning of 167 patients with TAK and 663 healthy controls using Illumina Human-Exome arrays to identify a replicating association between TAK and frequency of SNPs in the IL12B region. As a result, we found that the IL12B region on chromosome 5 (overall $p=1.7\times10^{-11}$, OR: 1.75, 95% CI: 1.42-2.16) and the MLX region on chromosome 17 (overall $p=5.2\times10^{-10}$, OR: 1.50, 95% CI: 1.28-1.76) as well as the SNP in the IL12B region (overall $p=2.4\times10^{-4}$, OR: 2.44, 95% CI: 1.89-3.12) exhibited significant associations. A significant synergistic effect of the SNP in the IL12B region and HLA-B*52:01 was found with a relative excess risk of 3.45, attributable proportion of 0.58, and synergy index of 3.24 (p=0.0028) in addition to a suggestive synergistic effect between the SNP in the MLX region and HLA-B*52:01 (p=0.027). We also found that the SNP in the IL12B region showed a significant association with clinical manifestations of TAK, including increased risk and severity of aortic regurgitation, a representative severe complication of TAK. Detection of these susceptibility loci in the current study will provide a new insight to the basic mechanisms of TAK pathogenesis. Our findings indicate that IL12B plays a fundamental role on the pathophysiology of TAK in combination with HLA-B*52:01 and that common autoimmune mechanisms underlie the pathology of TAK and other autoimmune disorders such as psoriasis and inflammatory bowel diseases in which IL12B is involved as a genetic predisposing factor.

888W

Identification of CNVs association in saliva flow using PennCNV. M. Lee1, K. T.Cuenco1, X. Zheng1, E. Feingold1, D.E. Weeks1, R.J. Weyant1, R.J. Crout1, D.W. McNeil2, M.L. Marazita1. 1) University of Pittsburgh, Pittsburgh, PA; 2) West Virginia University, Morgantown, WV.

CNVs are a complex trait also impacted by low saliva flow which helps establish the local environment that interacts with the teeth. The estimated heritability of saliva flow is 49% in COHRA subjects. Past GENEVA GWAS of saliva flow have not considered the potential impact of copy number variations (CNVs) as part of the genetic variability explaining saliva flow differences. The purpose of this study is to identify CNVs that are associated with saliva flow in the COHRA study population and to conduct a comprehensive investigation of these genetic variants. Methods: 1506 Caucasian subjects of which 75 years were enrolled in the study. The CNV calls were generated using the PennCNV software. Poor-quality samples identified with log R ratio standard deviation greater than 0.3 are excluded. CNVs with copy number $<2$ were defined as deletions, and those with copy number $>2$ as duplications. Genotyping data analyses with saliva flow are being conducted. We also compare CNV association results with previous hits from single SNP GWAS to check long-range linkage disequilibrium between genome regions. Results: We are currently preparing CNV calls and their association with saliva flow. Identified associations of CNVs with saliva flow will enrich understanding of genetic variability in caries-related traits. Support: DE020127 (K.T.Cuenco), and DE018903 and DE014899 (ML Marazita).
Atrioventricular septal defects (AVSD), a severe congenital heart defect (CHD), occur in the general population in ~1 in 10,000 births. Nearly 20% of infants with Down Syndrome (DS) have an AVSD, representing a ca. 2000-fold increased risk compared to the euploid population. We hypothesize that in the presence of an extra chromosome 21, otherwise benign copy number variants (CNVs) act in an additive manner to explain the increased penetrance of DS-associated AVSD. We have used the Affymetrix SNP 6.0 genotyping platform to comprehensively characterize CNVs in 538 DS samples, consisting of 243 cases (DS + complete AVSD) and 295 controls (DS - CHD). We implemented strict quality control filters to minimize false positive calls, including 3 algorithms to make the CNV calls (BEAST, GADA and GLAD). We also required each putative CNV call to contain > 10 SNPs within the interval. After excluding CNVs overlapping centromeres, we identified 1,711 total deletions (781 in cases and 930 in controls) and 1,302 duplications (607 in cases and 695 in controls). Association analyses in PLINK uncovered a ~275 kb deletion overlapping the MIR1324 gene on chr 3, which after permutation correction for multiple testing is genome-wide significant (p<0.049). Additionally, we have identified large rare duplications overlapping the candidate genes POSTN, MYH11, CTSB and PLCB1, although standard burden tests in PLINK did not identify statistically significant differences between cases and controls in the number or size of the CNVs. This analysis suggests that the genetic factors contributing to the increased prevalence of DS-associated AVSD are complex and heterogeneous. Our efforts are currently focused on the analyses of rare variants in the data set that might contribute to AVSD susceptibility in this sensitized population.
888W

Identification of SPOCK2 as a susceptibility gene for bronchopulmonary dysplasia. A. Hadchouel1, 2, 3, X. Durrmeyer4, E. Bouzigon5, 6, 7, R. Incitti5, J. Huuskko8, P.H. Jarreau10, R. Lenclen11, F. Demenais5, 6, 7, K.M.L. Franco-Montoya2, I. Layouni2, J. Patkai10, J. Bourbon5, M. Hallman9, C. Danan8, 12, C. Delacourt1, 2, 3, 1) Pneumologie Pédiatrique, Hôpital Necker Enfants Malades, APHP, Paris, France; 2) Université Paris Descartes - Sorbonne Paris Cité, Institut Imagine; 3) INSERM, U955, Créteil, France; 4) Réanimation Néonatale, Centre Hospitalier Intercommunal, Créteil, France; 5) INSERM, U946, Paris, France; 6) Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d’Hématologie, Paris, France; 7) Fondation Jean Daussel-Centre d’Etude du Polymorphisme Humain, Paris, France; 8) Institut Mondor de Recherche Biomédicale, INSERM, U955, Plateforme Microarray, Faculté de Médecine, Créteil, France; 9) Institute of Clinical Medicine, Department of Pediatrics, University of Oulu, Oulu, Finland; 10) Service de Réanimation Néonatérale, CHU Cochin-Port Royal, Paris, France; 11) Service de Réanimation Néonatérale, CH Poissy-Saint Germain, Poissy, France; 12) Unité Fonctionnelle de Recherche Clinique, Centre Hospitalier Intercommunal, Créteil, France.

Bronchopulmonary dysplasia (BPD), defined as a requirement for oxygen supplementation at 36 weeks of postmenstrual age (PMA), is the most common chronic respiratory disease in premature infants and its treatment places major demands on health services. Despite considerable advances in the care of very-low-birth-weight (VLBW) infants, BPD still occurs among 20-40% of survivors. Besides the recognized detrimental effects of environmental factors, VLBW twin concordance studies suggested a role of genetic factors. Candidate-gene studies failed to identify robust associations in this setting. The objective of our work was to perform a first genome-wide association studies (GWAS) in BPD. We prospectively evaluated 418 VLBW neonates (PMA lower than 28 weeks), of whom 22% developed BPD. Two discovery series were created using a DNA pooling strategy in neonates from white and African ancestries. Association between single nucleotide polymorphisms (SNPs) and BPD were investigated by two different analysis methods. SNPs associated with the disease were confirmed in an independent replication population. Selected genes were then explored by fine mapping and associations were replicated in an external Finnish population of 213 newborns.

889T


Background: Diabetic retinopathy (DR) is a major health problem affecting most patients with diabetes, yet only a subset progress to severe DR (SDR). Duration of diabetes and glycemia are the main risk factors for DR. Several lines of evidence suggest a genetic contribution to the risk of SDR; however, no genetic variant has shown convincing association with DR in genome-wide association studies (GWAS). Purpose: To identify common polymorphisms associated with SDR. Methods: White persons with type 1 diabetes (T1D) participating in the Epidemiology of Diabetes Interventions and Complications (EDIC, n=1904) and Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR, n=603) were genotyped by illumina Human1M and Omni1-Quad assays, respectively. Following quality control procedures, genotypes were coded and analyzed on the basis of 2.5M autosomal SNPs. Results: Analysis of SDR (n=328 SDR+; n=1555 SDR-) in T1D participants from the EDIC and WESDR cohorts (primary cohort, secondary cohort by treatment group - conventional or intensive) was performed by conditional logistic regression with adjustment for covariates in the association analysis of SDR with additive random genotype. The association of top SNPs was evaluated in three independent white T1D cohorts: Genesis-GeneDiab (n=502), Steno (n=936), FinnDiane (n=2194). In meta-analysis, each copy of the risk allele of any of the top SNPs was associated with 2.98 (OR=2.22, 95% CI, 1.66-2.98). Although, the top SNPs did not show significant association with SDR in the three replication studies (OR=2.22, 95% CI, 1.66-2.98). The direction of effect remained consistent in all but one of the examined populations. DPP10 is strongly expressed in the brain and pancreas and alters the expression and biophysical activity of voltage gated potassium channels. It has been associated with asthma by linkage/association mapping. Unlike DPP4, target for anti-diabetic gliptins, DPP10 does not show protease activity. Conclusion: DPP10 is the first locus to show strong evidence for association with SDR in a meta-GWAS.
We created a weighted GRS in the Osteoarthritis Initiative (OAI) based on greater rKOA presence if BMI were indeed a causal risk factor. Methods: We applied genome-wide complex trait analysis (GCTA) to the WTCCC-T2D GWAS dataset representing subjects from the United Kingdom (1924 cases, 2938 controls). We estimated the proportion of phenotypic variance attributable to additive effects of all variants interrogated in these GWAS (i.e. chip-based heritability), as well as from a much smaller set of variants identified as eQTLs for human skeletal muscle and adipose tissue. The estimate of chip-based heritability explained by the total set of GWAS variants is 60% (SE = 6%). The estimates of heritability explained by SNP subsets enriched for eQTLs mapped in human adipose and skeletal muscle tissue are 21% (SE = 7%) and 22% (SE = 8%), respectively. The combined set of adipose and skeletal muscle eQTL-enriched SNP subsets explained 33% (SE = 7%) p-value = 4×10^-6) of the phenotypic variation whereas the complement set of genome-wide SNPs explained less heritability (30%, SE = 7%), despite representing 3X more variants. The heritability estimate corresponding to the combined eQTL-enriched set was also greater and more statistically significant than the estimate of 6% (SE = 5%, p-value = 0.1) corresponding to a SNP set enriched for eQTLs mapped in human lymphoblastoid cell lines (LCLs) despite there being a greater number of SNPs (8% more) represented in the LCL eQTL-enriched set. Taken together, these results support our hypothesis that common eQTLs mapped in insulin-regulatory tissues account for a substantial portion of the variance in liability to T2D.

**891W**

HIGHER BMI GENETIC RISK SCORE IS ASSOCIATED WITH PRESENCE OF RADIOGRAPHIC KNEE OSTEOARTHRITIS BUT NOT PROGRESSION IN THE OSTEOPATHY INITIATIVE STUDY. M.S. Yau1, R.D. Jackson2, M.C. Hochberg3, S. Krishnan3, D.J. Duggan4, B.D. Mitchell4, L.M. Yerges-Armstrong4, 1) University of Maryland, Baltimore, MD; 2) Ohio State University, Columbus, OH; 3) Translational Genomic Research Institute, Phoenix, AZ.

**Purpose:** Increased BMI is highly associated with knee osteoarthritis, but the association could be a cause or effect of the condition. To distinguish between these alternatives, we assessed the association of BMI genetic risk score (GRS) with radiographic knee osteoarthritis (rKOA) and structural progression, reasoning that higher BMI GRS would be associated with greater rKOA presence if BMI were indeed a causal risk factor. **Methods:** We created a weighted GRS in the Osteoarthritis Initiative (OAI) based on published effect sizes for 32 robustly associated BMI risk alleles (P<3×10^-8) and tested its association with KOA presence and progression using logistic regression, adjusting for sex and age. The OAI is a longitudinal cohort of racially diverse high-risk individuals ages 45-79 years at baseline. The current analysis is restricted to Caucasian participants with at least two consensuses radiographs over four years. Genome-wide genotyping was conducted on the Illumina 2.5M platform and imputed to the 1000 genomes (CEU/June2011) using Minimac. We defined rKOA cases as individuals who have a KL grade ≥ 2 (n=1,912) in one or more knees and rKOA controls as individuals who have a KL grade ≤ 1 in both knees (n=1,225). Progressive cases have an increase in KL grade from baseline KL grade=1, 2, or 3 in one or more knees or total joint replacement (n=608). Controls for progression have no change in KL grade from baseline KL grade=1, 2, or 3 in both knees (n=1,238). **Results:** BMI GRS was significantly associated with presence of rKOA (adjusted OR=1.22, 95% CI=1.07-1.40), but not with rKOA progression (adjusted OR=1.08, 95% CI=0.90-1.29). Adjusting for baseline BMI substantially attenuated the association between GRS alleles on rKOA (adjusted OR=1.13, 95% CI=0.98-1.30), further supporting a causal association between overweight/obesity and rKOA. Also, FTO rs1558802 alone was associated with rKOA (adjusted OR=1.37, 95% CI=1.05-1.79). **Conclusions:** Higher BMI GRS is associated with rKOA, but not rKOA progression, even without the most robustly associated FTO allele. This supports the hypothesis that the effect of BMI precedes disease onset and plays a role in the onset of rKOA, but not in the trajectory of the disease. These results may have practical implications for management of rKOA.

(1) Laboratory for Genotyping Development, Center for Integrative Medical Science, RIKEN, Yokohama, Japan; (2) Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; (3) Laboratory for Statistical Analysis, Center for Integrative Medical Science, RIKEN, Tokyo, Japan; (4) Laboratory for Medical Science, Center for Integrative Medical Science, RIKEN, Yokohama, Japan; (5) Department of Medicine, Division of Gastroenterology, Social Insurance Chuo General Hospital, Tokyo, Japan; (6) Department of Internal Medicine, Faculty of Medicine, Toho University, Chiba, Japan; (7) Department of Gastroenterology, Sapporo Kosei Hospital, Sapporo, Japan; (8) First Department of Internal Medicine, Sapporo Medical University School of Medicine, Sapporo, Japan; (9) Department of Gastroenterology, Fukuoka University Chikushi Hospital, Fukuoka, Japan; (10) Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

Crohn’s disease is (CD) an inflammatory bowel disease (IBD) induced by multiple genetic and environmental factors. The prevalence of IBDs is much lower in Asian countries, including Japan, than in Western countries, but it is rapidly increasing. Genome-wide association studies (GWAS) have identified many genetic factors for CD in the European population, but information in other ethnic groups is scarce. Recently, we reported two novel genetic loci for CD in a Japanese population. To search for additional candidate loci, we performed genome-wide genotype imputation in the GWAS cohort (372 cases and 3,366 controls) using 286 East-Asian subjects from the 1000 Genomes Project March 2012 release (phase 1, version 3). From analysis of the imputed GWAS data, we found 44 additional candidate lead SNPs representing putative novel susceptibility loci for CD. To conduct tests of association at each marker, GEMMA software was used. Two SNPs were identified that exceeded the threshold for significance (p < 1.8e-7) after adjusting for multiple testing. One SNP (rs12491628; p = 6.9e-12; OR = 2.77) resides on 3p24.1, approximately 100 Kb upstream of the LRRC35 gene. The other SNP (rs11894081; p = 1.3e-8; OR = 2.56) is located on 1p36.24 in the TNFRSF9 gene. Discussion We have identified two potential genetic variants associated with the development of rotator cuff tears. To our knowledge, these findings represent the first attempt to identify genetic loci influencing the development of rotator cuff tears in a genome-wide association study. Further research is being conducted to further characterize these regions with the goal of identifying specific genes associated with tearing. References. 1. Tashiz RJ, Fanham JM, Albright FS, Teerlink CC, Cannon-Albright LA. J Bone Joint Surg Am 2009; 91:1136-42.

Evaluation of the common genetic architecture of problematic peer relationships. B. St Pourcain1,2, C.M.A. Haworth1, O.S.P. Davis1, N.J. Timpson1, T.A. MacWinnan1, J. Kemp1, D. Evans1, S.M. Ring1, W. McArdle1, J. Golding1,5, R. Plomin2, G. Davey Smith1. 1) MRC CAiTE, University of Bristol, Bristol, United Kingdom; 2) School of Oral and Dental Sciences, University of Bristol, UK; 3) MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, UK; 4) School of Social and Community Medicine, University of Bristol, UK; 5) Centre for Child and Adolescent Health, University of Bristol, UK.

Peer interaction plays an important role in the development of social competence, and problematic childhood peer relationships often persist into later maladjustment. Some links between early peer rejection and later maladaptive functioning however might be mediated through an underlying pathology. For example, impairments in social interaction skills are characteristic and heritable symptoms of the autistic dimension. Our study utilizes a twin study and a genome-wide analysis, to investigate genetic influences contributing to impaired peer relationships during childhood and adolescence. Heritability was estimated using 7,836 UK twin pairs (TEDS) with parent-report on peer problems at 4, 7, 9 and 11 years. Using a linear ACE model, we found that a considerable proportion of the phenotypic variance is attributable to genetic factors (0.60 < h² < 0.71), and that some genes affecting peer problems during development remain the same (0.32 < r_e < 0.69). Across development, we observed primarily non-shared environment influences (0.27 < e² < 0.32), with minimal evidence for shared environmental effects (0.02 < c² < 0.09). A subsequent genome-wide screen was conducted in ≤8000 children from a large UK birth-cohort (ALSPAC) with parent-report on peer problem at 4, 7, 8, 10, 12, 13 and 17 years. The association of shared and allele-dose effects was assessed using a Poisson family model, and carried out for each time-point individually, thus allowing for genetic heterogeneity. The strongest signals (2×10^-1 < P < 1x10^-7) were eventually modelled longitudinally and further investigated in 793 autism pedigrees (AGRE). Two population-based signals contributed to risk for autism and were consistent in their direction of effect. However, combined evidence (based on P-values) only reached suggestive levels of significance. This included an age-independent SNP in the vicinity of GIL13 at 9p24.2 (P=0.0030, Fisher-Combined-P=2.8x10^-7) and an age-dependent SNP effect (Age-SNP-P_{ALS-PAC}=4.9x10^-7) near GHR at 5p12, which was stronger in later adolescence (P_{ALS-PAC}=2.5x10^-10, P_{GHR}=0.00058, Fisher-Combined-P=2.7x10^-10). Signals are currently being followed-up in ≤2835 independent individuals in TEDS (age 4-11 years). Together, our findings support the contribution of common genetic effects to variation in problematic peer relationships, though many more samples might be required to reliably identify individual SNP signals.
Reclassification in genetic risk prediction over time. R. Barfield\textsuperscript{1}, J. Krier\textsuperscript{1}, R. Green\textsuperscript{2}, P. Kraft\textsuperscript{1,3}. 1) Department of Bioinformatics, Harvard School of Public Health, Boston, Massachusetts; 2) Harvard Medical School Genetics Training Program, Boston, Massachusetts; 3) Division of Genetics, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts; 4) Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts.

Genome-wide association studies have identified hundreds of common alleles with modest effects on risk of complex disease. The clinical utility of genetic risk profiles based on these variants depends crucially on the number and effect size of identified loci, and how stable the predicted risks are as additional loci are discovered. Individuals flagged as high risk at one time may be reclassified as low risk or vice versa as more loci are identified. Motivated by the development of a risk report for common complex disease as part of the MedSeq project on the integration of whole genome sequencing into clinical medicine, we quantify this reclassification using breast cancer (BrCa) and heart disease (CHD) as examples. Published results for potential causal SNPs were taken from the NHGRI catalogue. We simulated genotypes for a EUR population (n=100000) for these SNPs from the 1000 genomes projects. We calculated the predicted odds for disease for an individual as the product of the genotype specific odds ratios across all the risk markers known in 2007, 2009, 2011 and 2013. The odds were then normalized by the mean odds for all members of the simulated cohort for that year. The range of the predicted risks increased from 2007 to 2013: the 95th risk percentiles rose from 1.34 and 1.33 times the population average to 2.11 and 2.22 for CHD and BrCa, respectively. This caused 6% of the population to be reclassified from lower to 2x average CHD risk in 2007 to higher in 2013. The reclassification proportion for BrCa was 7%. The proportion of subjects at very high risk (greater 5x population average) increased from 2007 to 2013 (0 to 4x10^-4 and 7x10^-4 for CHD and BrCa respectively). We used the distribution of GWAS-identified risk markers to estimate the number of as-yet-unknown common risk alleles and their effect distribution, and used this to project the number of risk markers identified in future studies. The future reclassification from above 2x risk to below (or vice versa) projected from doubling available GWAS samples was notably smaller. This suggests that the reclassification has already occurred and has been incorporated into risk prediction tools. This study is useful for understanding the impact of reclassification on the use of genetic risk information in clinical practice and the way that risk prediction tools should be used in the future.

89T

Genome-wide association study identifies novel loci associated with abdominal obesity in Africans. A. Doumaly\textsuperscript{1}, A. Adeyemo\textsuperscript{1}, G. Chen\textsuperscript{1}, P. Nkrumah\textsuperscript{4}, A. Berendts\textsuperscript{1}, M. Zhou\textsuperscript{1}, H. Huang\textsuperscript{1}, D. Shriner\textsuperscript{1}, D. Ngare\textsuperscript{1}, O. Fonsannamide\textsuperscript{2}, T. Johnson\textsuperscript{1}, J. Oli\textsuperscript{1}, G. Okafar\textsuperscript{2}, B. Eghan\textsuperscript{5}, K. Agyenim-Boatang\textsuperscript{1}, J. Adeleye\textsuperscript{1}, W. Balogun\textsuperscript{1}, C. Adebamowo\textsuperscript{2,3,4}, A. Amoah\textsuperscript{5}, J. Accheampong\textsuperscript{1}, C. Rotimi\textsuperscript{1,6}. 1) NHGRI/CRGGH, National Institutes of Health, Bethesda, MD; 2) Department of Medicine, University of Lagos, Lagos, Nigeria; 3) Department of Medicine, University of Nigeria Teaching Hospital, Enugu, Nigeria; 4) Department of Medicine, University of Science and Technology, Kumasi, Ghana; 5) Department of Medicine, University of Ibadan, Ibadan, Nigeria; 6) Department of Medicine and Therapeutics, University of Ghana Medical School, Accra, Ghana; 7) Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, MD; 8) Department of Mental Health and Behavioral Sciences, M.U. School of Medicine, Eldoret, Kenya.

Obesity, especially abdominal obesity (AOb), is a risk factor for cardiovascular diseases and type 2 diabetes (T2D). AOb is heritable and several loci have been shown to influence its distribution at the population level. However, most prior studies of genes influencing AOb were performed in populations of non-African ancestry, a limitation given known differences in AOb by ethnicity. In this study, we conducted the first genome-wide association study of AOb in Africans. The study participants comprised 1808 Africans enrolled in the Africa America Diabetes Mellitus (AADM) study. Waist circumference (WC) and Waist-to-hip ratio (WHR) were used as measures of abdominal adiposity. Genotyping was carried out using the Affymetrix Axiom PANAFR SNP array (~2.1 million SNPs) imputed using the latest 1000 Genomes cosmopolitan reference for a total of 15 million SNPs included in the analysis. Association tests were adjusted for age, gender, T2D, BMI, and the first 3 principal components (PCs) of the genotypes. A cluster of SNPs on chromosome 2 (2p25.3, near FAM110C) and rs116158389 on chromosome 19 (19p13.2, near RDH8) were associated with WC at genome-wide significance level (p<5x10^-8). The top-ranked SNP on chromosome 2 (rs111441856, FAM110C, p=5.1x10^-10) is in moderate- to complete LD (0.64<r2<1) with the other associated SNPs in the region, indicating the likelihood of a single signal at the locus. FAM110C is expressed in adipocytes and fuel metabolism. FAM110C has been implicated in metabolic processes including mitochondrial remodeling and Akt1 activation. The associated chromosome 2 variants are polymorphic in sub-Saharan Africans but are monomorphic in other continental populations. Several loci were nominally (7.6x10^-6<p-value<1.6x10^-8) associated with WHR including variants on chromosomes 19 (near RDH8, ZNF814, and MIR5589), 1 (RALGPS2), 2 (near REG3G), and 1.6x10^-6) associated with WHR, including variants on chromosomes 19 (near RDH8, ZNF814, and MIR5589), 1 (RALGPS2), 2 (near REG3G), and 11 (in MPM3). This first GWAS of AOb in Africans identified novel loci that may provide new insights into the genetic architecture of obesity. While additional studies are needed to confirm and extend the findings, these results underscore the importance of studying multiple global populations in genetic studies of complex traits.

89F

Genome-wide association analysis shows the highly polygenic character of age-related hearing impairment. E. Fransen\textsuperscript{1,2}, S. Bonneux\textsuperscript{1}, J.J. Corneveaux\textsuperscript{1}, I. Schrauwen\textsuperscript{1,3}, F. di Bernardo\textsuperscript{1}, C.H. White\textsuperscript{4}, J.D. Omen\textsuperscript{5}, P. Van de Heyning\textsuperscript{6}, U. Ambrosieth\textsuperscript{1}, M.J. Huetelema\textsuperscript{1}, G. Van Camp\textsuperscript{1}, R.A. Friedman\textsuperscript{1}. 1) Center Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) StatUa center for statistics, University of Antwerp, B-2000 Antwerp; 3) Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, Arizona 85004 USA; 4) Audiology Unit, Dept. of Clinical Sciences & Community Health, University of Milan; Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, Milan, Italy; 5) Cell Biology and Genetics Division, House Research Institute, Los Angeles, CA 90057; 6) Department of Otolaryngology, University Hospital of Antwerp, B-2650 Edegem, Belgium.

We have performed a genome-wide association study (GWAS) to identify the genes responsible for age-related hearing impairment (ARHI), the most common form of hearing impairment in the elderly. Analysis of common variants, with and without adjustment for stratification and environmental covariates, revealed no rare variants as significant enrichment. Therefore, analysis showed no variants with genome-wide significance. No evidence for replication of any previously reported genes was found. The phenotype depends on the aggregated effect of a large number of SNPs, of which the individual effects are undetectable in a modestly powered GWAS. We estimate that 22% of the variance in our dataset can be explained by the collective effect of all genotyped SNPs. A score analysis showed a modest enrichment in causative SNPs among the SNPs with a p-value below 0.01. This is in line with the idea that ARHI is highly polygenic in nature, with probably no major genes involved.
900W
Metabolic pathways in relation to obesity: Untargeted metabolomic profiling in a large population-based study. J. Kumar1, A. Ganna1, T. Fall1, J. Prenni2, C. Broekkling1, J. Prince1, L. Lind1, E. Ingelsson1, 1) Department of Medical Sciences, Molecular Epidemiology, Uppsala University and Science for Life Laboratory, Uppsala, Sweden; 2) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 3) Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, U.S.A; 4) Department of Medical Sciences, Uppsala University, Uppsala, Sweden.

Introduction: Obesity is the main cause of various disorders such as type 2 diabetes, and cardiovascular diseases and has become a serious public health problem of increasing prevalence. In the present study, we investigated the associations of body mass index (BMI) with circulating metabolites in a non-targeted metabolomics experiment and then performed a genome-wide association study (GWAS) of associated metabolites. Materials and Methods: Ultra-performance liquid chromatography coupled with tandem mass spectrometry was employed for performing a non-targeted metabolite profiling in two population-based cohort studies (TwinGene; N=1520 and PIVUS, N=970). Linear regression analyses were performed to test the association of BMI with different metabolites adjusting for potential confounding factors (age, sex, seasonal variation, storage, handling time and time since last meal). Significant findings were selected by controlling the false discovery rate (FDR) < 5%. GWAS was performed on the significant metabolite features to identify their associated loci. Results: A total of 8,185 and 11,056 molecular features were observed after quality control in PIVUS and TwinGene, respectively. In total, 2,281 molecular features were found to be significantly associated with BMI in PIVUS (FDR<1%) and 1,216 of these features were also found to be associated with BMI in TwinGene (FDR<1%). Most of the significant features represent 113 different metabolites (based on similar retention time and high correlation) of which seven could be annotated with high confidence (1-oxy-2-hydroxy-sn-glycerol-3-phosphocholine, Glycerophosphocholine, negatively associated with BMI; Alpha-Linolenic acid, Caffeine, Dexoycholic acid glycine conjugate, L-Carnitine and L-Tyrosine, positively associated with BMI) utilizing a private compound library. GWAS of these seven metabolites from PIVUS and TWINGene showed that SNPs from carminine gene (rs1171617, p=2.41×10−6) is the significant association for circulating levels of circulating carnitine. Conclusion: In conclusion, the application of non-targeted metabolite profiling to evaluate two large population-based cohorts revealed 113 metabolites to be significantly associated with BMI. Seven of these metabolites were found to be significantly associated with circulating levels of L-Carnitine. Further efforts to improve annotation of the remaining are underway through additional experiments.

901T
Genomewide Association Studies of Lipids in Samoans. R.L. Minster1, N.L. Hawley2, G. Sun3, H. Cheng3, S. Viall3, R. Dekat3, D.E. Weeks4, S.T. McGarvey2, 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Weight Control and Diabetes Research Center, the Miriam Hospital, Providence, RI, USA; 3) The Albert Medical School, Brown University, Providence, RI, USA; 4) Department of Environmental Health, School of Medicine, University of Cincinnati, Cincinnati, OH; 5) Medical Specialist Clinic and National Health Services, Government of Samoa, Apia, Samoa; 6) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 7) Department of Epidemiology, International Health Institute, School of Public Health, Brown University, Providence, RI.

Genomewide association studies (GWAS) of quantitative traits in populations isolates affords a unique opportunity to identify quantitative trait loci. Studies in populations like Samoans can detect genes that are novel and population-specific as well as those that affect general human health. Obesity has greater prevalence among Samoans compared to many other ancestry groups, and identifying variation that affects lipid levels against this background could reveal novel mechanisms for intervention. Here we report the results of GWAS of levels of total cholesterol, high-density lipoprotein, low-density lipoprotein and triglycerides in 3,122 adults from Samoa, age, age2, sex, BMI, smoking and alcohol use were considered as covariates for each of the four phenotypes. A total of 871,188 autosomal single-nucleotide polymorphisms (SNPs) were genotyped and extensive quality control of the genotypes was performed. Five loci were identified with genomewide significance (p < 5 × 10−8), all of which have been reported previously (APOE, ZNF259, CETP, GCKR and LIPC). Six loci, three of them novel, were identified with suggestive associations between p = 10−6 and p = 5 × 10−8. An additional three loci had suggestive associations with p values between 10−6 and 10−8 with one or more of the four lipid traits. Additional studies in related Polynesian populations are necessary to verify these new findings. This work was supported by U.S. National Institutes of Health grant R01-HL093093 (P.I. Stephen McGarvey).

902F
Identification of genetic factors underlying asthma age-of-onset sub-phenotypes. C. Sarowski1,2, M.-H. Dizier3, I. Ahmed4, P. Margaritelou1, J. Coresh5, A.R. Folsom6, J.-M. Lahousse7, F. Demenais1,2, E. Bouzigon1,3, the EGEA cooperative group. 1) UMR946, INSERM, PARIS, France; 2) Univ. Paris Sud, Paris, France; 3) Univ. Paris Diderot, Paris, France; 4) CESP, INSERM, UMRS 1018, Villejuif, France; 5) McGill Univ., Montréal, Canada; 6) CNGEAA, Evry, France.

Asthma is a heterogeneous disease with variable clinical expression over the life span. The disease age of onset is one of the simplest features that can be used to differentiate asthma phenotypes. To characterize the genetic factors influencing asthma in age-of-onset specific manner, we conducted a GWAS using a multinomial regression model applied to 750 asthmatics categorized according to their age-of-onset and 1,085 non-asthmatics from the French EGEA study with HapMap2 imputed data. Asthmas were split into four age-of-onset sub-phenotypes: A) age-of-onset ≤4yrs (early-onset), B) 5-12yrs (before puberty), C) 13-20yrs (between puberty and adulthood) and D) > 20yrs (adult-onset). First, we applied an association test allowing heterogeneity of SNP effect between sub-phenotypes (Morris et al. Genet Epidemiol 2010) and detected 60 SNPs with P-value ≤10−5. Then, we tested whether these SNPs had a heterogeneous effect among the four sub-phenotypes. We identified 53 SNPs located in 16 regions with an interclass heterogeneity P-value ≤10−5. Among these regions, six had interclass association P-values ≤10−4. We confirmed the specific association between 17q12-q21 genetic variants and early-onset asthma (P = 10−6) (Bouzigon et al. N Engl J Med 2008). We also detected five new regions among which four loci with SNP effect restricted to one asthma age-of-onset sub-phenotype: 3q34 with phenotype A (P=5×10−7), 3q25 with phenotype B (P=2×10−6), 1q42.1 with phenotype C (P=3×10−5), and 2q15 with phenotype D (P=1×10−4). This analysis will be extended to GABRIEL Asthma consortium dataset. Thus, taking into account the age of onset in a multinomial regression model was a powerful tool to identify new genetic loci underlying complex diseases. Funded by: Région Ile de France, Fonds de Dotation "Recherche en Santé Respiratoire" & ANR-GEWIS-AM, GABRIEL.

903W
Genetic Association of Serum Magnesium Levels in African Americans - the Atherosclerosis Risk in Communities (ARIC) Study. A. Tinn1, A.R. Folsom1, N. Maruthur2, C.A. Friedrich3, J. Corelli3, E. Boenirmware4, W.H. Kao1, 1) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) University of Minnesota School of Public Health, Minneapolis, MN; 3) Johns Hopkins University School of Medicine, Baltimore, MD; 4) University of Mississippi Medical Center, Jackson, MS; 5) University of Texas School of Public Health, Houston, TX.

Background. Magnesium (Mg) is important in enzymatic reactions and the regulation of mitochondrial function and vascular tone. Low serum Mg is associated with a greater burden of cardiovascular disease, hypertension, and diabetes. African Americans (AAs) have both lower serum Mg and dietary Mg intake versus European Americans (EAs). Genome-wide association studies (GWAS) in EA identified six loci associated with serum Mg, but no GWAS of serum Mg have been conducted in AAs. Methods. We conducted the first GWAS of serum Mg in 2758 AA ARIC participants using nominal regression framework can be a powerful approach to identify new loci underlying complex diseases. Funded by: Région Ile de France, Fonds de Dotation "Recherche en Santé Respiratoire" & ANR-GEWIS-AM, GABRIEL.
904T

Genome-wide search for age- and sex-dependent genetic loci for human anthropometric traits: Methods and results from genome-wide meta-analyses across 310,000 individuals. T.W. Winkler1,2, Z. Kutalik3,4, M. Graff5, A. Justice5, L. Barata5, M. Feitoza6, S. Chu7, R. Magi8, J. Czajkowsk9, T. Fall10, Y. Lu11, T.O. Kiplêlânen12, I.M. Hei11,13, I. Boreck8, K.E. North4, R.J.F. Loos1,14, GIAN'T (Genetic Investigation of Anthropometric Traits) Consortium. 1) Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; 2) Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland; 3) Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; 4) Institute of Social and Preventive Medicine, CHUV-UNIL, 1010 Lausanne, Switzerland; 5) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA; 6) Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA; 7) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, USA; 8) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 9) Division of Statistical Genomics, Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St. Louis, MO 63110, USA; 10) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 11) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 12) The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Copenhagen, Denmark; 13) Genetic Epidemiology, Helmholtz Zentrum Muenchen-German Research Center for Environmental Health, Neuherberg, Germany; 14) MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK3,4.

Height, adiposity and fat distribution differ between men and women and change over time (e.g. following menopause). Previously, genome-wide association meta-analyses (GWAMAs) of anthropometric traits revealed sexually dimorphic loci for waist-hip ratio, a measure of fat distribution, but little is known of whether genetic effects on anthropometric traits change with age, and whether such changes differ between men and women. To detect age and sex dependent genetic effects, we conducted GWAMAs stratified by age (50-year age groups) and sex (men>50y, women>50y). For each trait, we conducted inverse-variance weighted meta-analyses (one for each stratum), subsequently tested the pooled stratum-specific estimates for difference between the age-groups, between the sexes and for heterogeneity between the four strata and controlled each test at 5% FDR. To boost statistical power, we focused only on the SNPs that showed some overall association in the four strata combined (P<1e-5).

For BMI and WHR, our analysis yielded 13 loci with significant age-difference (near COBLL1, DDC, NEGR1, TN113K, SEC16B, TMEM18, ADCY3, AC016194.1, TCF7L2, STK33, FTO, MC4R and APOC1), of which 10 showed a stronger effect in the younger subgroup. For WHR, our analysis yielded 25 loci with significant age-difference (16 novel sexually dimorphic loci near MUC, GANAB, RPS6KA5, GORAB, MCT1, PLXND1, FAM13A, MAP3K1, NSD1, LY6F, NFE2L3, NKKX3-1, DANA10, KCN2, EDEM2 and EYA2; and all nine previously established sexually dimorphic variants for WHR). Of the 25 loci, 23 showed a stronger effect in women. Although we did not find any loci with significant sex-difference for BMI or age-difference for WHR, our scan on the heterogeneity between the four strata identified two additional loci for BMI (near CXXC5 and TSH2) and one additional locus for WHR (near BMP2), which might reflect potential age or sex dependency. Our results underscore the importance of sex- and age-stratified analyses to illustrate a sexually dimorphic and age-dependent genetic underpinning for anthropometric traits.

905F

Targeted resequencing of genome wide associated candidate regions for pediatric venous thrombosis. A. Witten1, A. Aming1, A. Barysenka1, Ch. Grote2, M. Hiersche3, F. Ruehle1, U. Nowak-Goett1, M. Stol1,1. Genetic Epidemiology of vascular disorders, Leibniz-Institute for Arteriosclerosis Research at the University of Münster, Münster, NRW, Germany; 2) Institute of Clinical Chemistry, University Hospital Schleswig-Holstein, Germany.

Venous thrombosis is a common multifactorial disease, which is influenced by environmental and genetic factors. Here we present a genome wide association study (GWAS) in 212 trios with pediatric venous thrombosis comprising affected children and their parents. For two SNPs exceeding the threshold for genome wide association (p<10-5) determined by 1.000.000 bootstrap permutations replication was conducted in 201 trios with thromboembolic stroke (TS). Among these, rs1304029 and a SNP curiously missing the threshold for permuted p-value (rs2748331) reside in a region on chromosome 6q13 comprising the gene for beta-1,3-glucoronyltransferase 2 (B3GAT2), and are associated with pediatric VTE (rs1304029; p=1.42x10^-6, rs2748331; p=6.11x10^-6). rs2748331 was replicated (p=0.00719) in our GWAS on pediatric TS (combined p=7.88x10^-7). Twenty-seventh additional SNPs were associated at confident permuted P-values (p<1x10^-4). To investigate possible causative genetic variants in the resulting linkage disequilibrium based candidate regions, we performed a next generation resequencing approach in 24 affected children and 24 unaffected siblings. The selected target regions of about 11 Mb in total comprise 30 gene regions from 16 chromosomes. Custom target enrichment was performed using the NimbleGen SeqCap EZ Choice technology. The resulting DNA libraries were paired-end sequenced (100 cycles) on an Illumina HiScanSQ instrument, generating a total of 12.5 Gb sequence data in total and 91.8% bases with a QScore > 30. Sequence reads were mapped by using the BWA algorithm and analyzed by GATK yielding in 83% median target specificity and median target region coverage of 196x. Variant annotation was done by using SNPEFF and Annovar software tools. A sibling TDT test was applied on the 41.478 identified variants, 4.062 of which were novel to compare the two sample groups. 23 significant (p<0.05) coding non-synonymous or UTR SNPs in 10 genes were identified and selected for subsequent validation in 258 individuals from the full cohort of 258 nuclear families. Future studies elucidating the functional relevance of these genetic variants are warranted to further elucidate the role of these in the pathogenesis of DVT.
906W

A meta-analysis of genome-wide association studies for adiponectin level in East Asians identifies a novel locus near WDR11-FGFR2. Y. Wu1, H. Gao2,4, H. Li4, Y. Tabara5, M. Nakatoch6, Y. Fujiwara6, E.J. Park10, F. Vaidhyanad1, M. Fogarty1, W. Wen11, X. Shu12, C. Shin12, S. Hje16, L.M. Chuang13,14, T. Miki15, M. Yokota16, X. Lin17, K.L. Mohlke18, E.S. Tai17,18, The Asian Genetic Epidemiology Network (AGEN) adiponectin working group. 1) Department of Genomics, University of North Carolina, Chapel Hill, NC; 2) Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore; 3) NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 5) Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; 6) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 7) Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan; 8) Division of Bioinformatics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Taiwan; 9) Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan; 10) Institute for Health Promotion and Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Republic of Korea; 11) Division of Epidemiology, Department of Medicine, National Taiwan University, Taipei, Taiwan; 12) Department of Internal Medicine, Korea University Ansan Hospital, Ansan, South Korea; 13) Division of Endocrinology and Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 14) Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan; 15) Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Ehime, Japan; 16) Department of Genome Science, Aichi-Gakuin University, Nagoya, Japan; 17) Department of Rheumatology, National Hospital Organization, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore; 18) Duke-NUS Graduate Medical School, National University of Singapore, Singapore.

Blood levels of adiponectin, an adipocyte-secreted protein correlated with metabolic and cardiovascular risk, are highly heritable. Genome-wide associa-
tion (GWA) studies have identified 14 loci harboring variants associated with adiponectin level. To identify novel adiponectin-associated loci, particularly those of importance in East Asians, we conducted a meta-analysis of GWA studies for adiponectin in 7,827 individuals, followed by in silico and de novo replication in 4,298 and 5,724 additional individuals of the Asian Genetic Epidemiology Network (AGEN). Of the adiponectin-associated loci previously reported, we confirmed the association at six loci including CHD13, ADIPOQ, PEPD, CMP, ZNF664 and GPR109A. We identified a novel signal on chromosome 10, −300 kb from WDR11 and −300 kb from FGFR2 (P = 1.9 × 10−14). We also provided suggestive evidence of another locus near OR26T-LALBA on chromosome 12 (P = 9.4 × 10−9). FGFR2 is a strong candidate gene implicated in adipocyte hyperplasia and hypertrophy; the novel signal at WDR11-FGFR2 explained 0.6% of the total variation in adiponectin. Despite a consistent direction of allelic effect, the index SNP at WDR11-FGFR2 did not show strong evidence of association (P = 0.035) in > 29,000 Europeans in the publicly released ADIPOGen data. The difference in the trait-increasing allele frequency (AGEN: 0.57; ADIPOGen: 0.24) and the significant level of association across populations suggested that a variant at WDR11-FGFR2 might have a larger genetic effect on adiponectin level in East Asians than Europeans, or the pairwise LD between the index SNP and the un-typed causal variant vary across different populations. The adiponectin-increasing allele of the index SNP exhibited evidence of association with a decreased BMI-adjusted waist-hip-ratio in 17,560 East Asians (P = 9.8 × 10−9). At WDR11-FGFR2, twenty-two candidate variants that span 65 kb are in moderate to high LD (r2 > 0.6) with the index SNP, and at least four candidate variants are located at or near enhancer marks in adipose tissue. These findings improve knowledge of the genetic basis of adiponectin variation, demonstrate the shared allelic architecture of adiponectin and central obesity, and motivate further studies of underlying biological mechanisms.

907T

Genome-wide association analysis of skeletal muscle fiber types. T. Karadeniz1, N. Oskolakov2, C. Latenvall2, S. Keldson3, A. Mahajan4, L. Lind2, E. Ingelsson1, L. Groop5, P. Franks6, A.P. Morris7, O. Hansson7, C.M. Lindgren1, 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden; 3) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 4) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 5) Department of Clinical Sciences, Genetic and Molecular Epidemiology, Lund University, Malmö, Sweden.

The human body contains over 400 skeletal muscles, together constituting 40-50% of the total body weight. Skeletal muscle is a key organ in many basic bodily functions, such as movement, respiration and thermogenesis. Individual muscles consist of a combination of 3 types of fibers: type I fibers (red, slow and oxidative), type 2A (red, fast and oxidative) and type 2X (white, fast and glycolytic). It has been previously shown that muscle fiber composition is associated with obesity, weight loss and type 2 diabetes. Therefore, it is important to explore the genetic make-up of muscle fiber composition to better understand the factors affecting the risk of such conditions. We conducted a genome-wide association study (GWAS) of these three skeletal muscle fiber types in a total of 656 individuals from three independent cohorts from Sweden. Genotype data in each cohort were imputed up to 35 million variants from the 1000 Genomes ‘all ancestries’ reference panel (March 2012). Within each cohort, phenotypes were inverse rank normalized and tested for association with each high-quality imputed variant (info>0.4) under an additive model after adjustment for age and body-mass index. After a conservative double genomic control correction, only 1 variant (rs74979762 (minor allele frequency (MAF)=0.005, p=8.04 × 10−4), which is located downstream of SIA1L2, within a potential transcription factor binding site. There were also three loci associated with fiber type 2X. The first is intronic (rs145631867, MAF=0.02, p=2.23 × 10−5) and maps to NYAP2 and the second locus is intronic (rs149081100, MAF=0.02, p=2.96 × 10−5) maps to ADRA1B, within a DNase I hypersensitivity region and a possible strong enhancer-binding site for glucocorticoid receptors functioning in anti-inflammatory immune responses and regulation of glucose metabolism. In conclusion, we have performed the first GWAS of detailed physiological measures of human skeletal muscle fibre. The genetic associations we observe suggest that the make-up of fiber types in skeletal muscle are under at least partly genetic control.
908F

Genome-wide association study of skin pigmentation and tanning in African Americans. K. Batali1,2, E. Shah1,2, R.A. Kittles3,4. 1) Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL; 2) Institute of Human Genetics, College of Medicine, University of Illinois at Chicago, Chicago, IL; 3) Section of Hematology/Oncology, Department of Medicine, University of Illinois at Chicago, Chicago, IL; 4) Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, IL.

Background: Skin pigmentation is a complex trait that is strongly controlled by genes and sunlight exposure. The identification of pigmentation genes is important for understanding of skin cancer susceptibility and vitamin D status. Genome-wide association studies (GWAS) in European, Asian, and African-European admixed Cape Verdan populations have identified multiple loci associated with pigmentation traits. Many of these GWAS loci are similar across populations while several are population specific. To date, there has not been any GWAS performed in African Americans (AAs) that attempt to identify loci associated with pigmentation traits. Here, we report results of the first GWAS of skin pigmentation and tanning ability in AAs.

Methods: Skin pigmentation (M-Index) of 215 AA men from Washington, DC was measured on subjects’ forehead and upper inner arm using reflectometer and samples were genotyped using Illumina Infinium 1M-Duo bead array. After quality control, 994,618 SNPs were available for analyses. We tested association for the inner-arm (constitutive) M-Index using linear regression model adjusting for age and first 3 PCs and for tanning ability (inner arm M-Index subtracted from forehead M-Index) adjusting for age and first 3 PCs.

Results: A cluster of SLC24A5 SNPs were significantly associated with M-Index and three of them reached genome wide significant P-value (P<5.02 x 10^-8). The strongest association was found for rs2675345 (P=2.90 x 10^-8). SNP, rs3828107, in NR5A2 upstream of FLJ43763 located in 6p25, about 200kb attenuated, but rs9392299 in SLC24A5 showed stronger association with tanning ability (P=2.16 x 10^-11). When UV exposure was included in the regression model, association between this SNP and tanning ability was attenuated, but rs9392299 in FLJ43763 located in 6p25, about 200kb upstream of IRF4, a gene that was associated with tanning ability and sun sensitivity in European descent populations showed marginal significance (P=5.26 x 10^-14).

Conclusion: We found that SLC24A5 SNP showed stronger association with M-Index than previously reported GWAS SNPs in South Asian and admixed Cape Verdan populations. We successfully confirmed previous GWAS findings of 6p25 SNPs and tanning ability in our AA population. Our study also provides evidence for a novel loci on chromosome 1 associated with tanning ability.

909W

The First Genome-wide Association Study of Serum Lipids among Africans. A.R. Bentley1, L. Thrush1, G.J. Chen1, T. Tekola-Ayele2, J. Johnson1, M. Oli3, G. Okafor4, B.A. Eghan5, J. Argyenim-Boateng4, J. Adeleye5, W. Balogun5, C. Adedamoyo5, A. Amoah6, J. Acheampong6, D. N'gans6, A. Adeyemo6, C.N. Rotimi1. 1) Center for Research in Genomics and Global Health, Natl Human Genome Research Institute, Bethesda, MD; 2) Department of Medicine, University of Lagos, Lagos, Nigeria; 3) Department of Medicine, University of Nigeria Teaching Hospital, Enugu, Nigeria; 4) Department of Medicine, University of Science and Technology, Kumasi, Ghana; 5) Department of Medicine, University College Hospital, Ibadan, Nigeria; 6) Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, MD; 7) Department of Medicine and Therapeutics, University of Ghana Medical School, Accra, Ghana; 8) Department of Mental Health and Behavioral Sciences, Moi University School of Medicine.

The burden of chronic disease in Africa is increasing rapidly, with sharper increases predicted, yet the genetic determinants underlying relevant traits among Africans are largely unstudied. We have conducted the first genome-wide association study of serum lipids in Africa. The study participants comprised 1808 primarily West Africans enrolled in the Africa America Diabetes Mellitus (AADM) Study, characterized for total cholesterol (TC), HDL-cholesterol (HDLc), LDL-cholesterol (LDLc), and triglycerides (TG) and genotyped on the Affymetrix Axiom PANAFR array (~2.1 million SNPs). Imputation using the 1000 Genomes cosmopolitan reference yielded ~15 million SNPs. Association models were adjusted for gender, age, BMI, type 2 diabetes, and the first 3 principal components of ancestry. The most statistically significant results were for variants in the APOE region with LDLc (−0.5 mg/dL, lowest p=5.5 × 10^{-17} [rs7412]). A larger effect size was observed for a nearby rare variant that has not been reported among non-African populations, rs192607279 (−4.5 mg/dL LDLc, p=1.5 × 10^{-10}). Variants in this region were also associated with lower TC (lowest p=4.4 × 10^{-10} [rs51679753]). Common LDLR variants were associated with −0.3 mg/dL lower LDLc. rs11590558 near HTR2A was associated with LDLc (0.6 mg/ dL, p=4.0 × 10^{-10}); this SNP is monomorphic among those of European ancestry. Although not reaching genome-wide significance, a lead association in the FTO gene for body mass index (BMI) was replicated. No TG association was observed. The large proportion (56%) of these results in any African ancestry population. This first African GWAS of serum lipids confirmed the relevance of some known lipids loci (APOE, LDLR, and CETP) in this population, as well as revealing some novel associations, particularly among variants not present among non-African ancestry populations. Replication attempts of known loci were successful for 25% of SNPs investigated. The large proportion of variants for which this is the first replication in an African ancestry population is consistent with the hypothesis that a reduced influence of environmental risk factors in Africans residing on the continent may simplify detection of genetic associations compared to Africans in the Diaspora with more Western lifestyles.
910T
1) Department of Electrical Engineering and Computer Science (Institut Montefiore), University of Liege, Liege, Belgium; 2) Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, United States.

Genetic epistasis likely underlies most complex traits, including gene expression, yet is very difficult to detect using standard approaches. When considering the interactive effects on expression Quantitative Trait Loci (eQTL), interactions between genes, scattered throughout the genome, are referred to as trans-epistasis. In this work we aimed at identifying transcripts whose expression is regulated by SNP-SNP interactions using Model-Based Multifactor Dimensionality Reduction (MB-MDR). This model-free approach for epistasis detection involves reducing a high-dimensional GxG space to GxG factor levels that either exhibit high, low evidence or no evidence at all for their association to gene intensities of interest. We applied this method to a dataset consisting of 19,451 CD4+ lymphocyte expression phenotypes and genotypes from ~516,000 SNPs from 174 Caucasian non-smoking subjects from CAMP asthma cohort. Univariate analysis highlighted 1,844 genome-wide significant cis-acting eQTL. With MB-MDR, each of these was tested for potential synergetic interaction with the remaining ~516,000 trans-SNPs, excluding cis-SNPs located in genetic range between 2 Mb upstream and downstream of the eQTL gene. Our preliminary results reveal evidence for the existence of epistatic interactions in the regulation of gene expression. Out of the total of 1,086,601 cis-SNP/trans-SNP pairs with marginal p-value ≤ 0.001 based on 999 permutations, the top 10 trans-genes were all related to signaling and cell-cell adhesion pathways (PTPRD, CDH13, CNTN5, LRP1B). For example, the top 2 cis-trans SNP pairs ranked on the basis of their associated MB-MDR statistic (shown in brackets) - rs622614/LRP1B. For example, the top 2 cis-trans SNP pairs ranked on the basis of their associated MB-MDR statistic (shown in brackets) - rs622614/LRP1B. For example, the top 2 cis-trans SNP pairs ranked on the basis of their associated MB-MDR statistic (shown in brackets) - rs622614/LRP1B. For example, the top 2 cis-trans SNP pairs ranked on the basis of their associated MB-MDR statistic (shown in brackets) - rs622614/LRP1B.

911F
Genome-wide association study (GWAS) of major depressive disorder in N=9300 Han Chinese women using low-pass sequencing data. T. Bispell1, R.E. Peterson1, Y. Li2, W. Kretzschmar2, F. Yang2, J.H. Maes2, A.H. Fanous2, R.T. Grojean2, D.Croteau-Chonka3, S.A. Bacanu2, K.S. Kendler3, J. Flint2, CONVERGE consortium. 1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, PO Box 980126, Richmond VA, 23228, USA; 2) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, Oxfordshire, United Kingdom; 3) Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, No. 600 South Waiping Road, Shanghai, P.R. China; 4) Bejing Genomics Institute, Floor 9 Complex Building, Beishan Industrial Zone, Yiantian District, Shenzhen 518083, P.R. China; 5) Huashan Hospital of Fudan University, No.12 Middle Wulumuqi Road, Shanghai, P.R. China; 6) CTSU, Richard Doll Building, Old Road Campus, University of Oxford, Headington, Oxford OX3 7LF, United Kingdom.

Background: Major depressive disorder (MDD) is a common, complex psychiatric disorder and a leading cause of disability worldwide. Although modestly heritable (~30-40%), a complex genetic architecture has hindered efforts to identify robustly associated genetic risk variants. We sought to evaluate the evidence for common genetic variation in the etiology of MDD in a large, ethnically homogenous Chinese sample.

Methods: Using single nucleotide polymorphism (SNP) data imputed from low-pass (1.2X) sequencing data, we tested for association between common SNPs and MDD in 4168 cases and 4614 controls from the CONVERGE (China, Oxford and VCU Experimental Research on Genetic Epidemiology) project. We estimated the proportion of variation in disease susceptibility captured by common SNPs using the Genome-wide Complex Trait Analysis (GCTA) utility and considered the predictive value of scores constructed from GWAS results.

Results: In our preliminary analysis of ~7.5M SNPs, no single variant attained genome-wide significance (5×10^-8). Based on an estimated prevalence of 12%, common SNPs accounted for an estimated 15.7% of the variance in MDD risk (95%CI=[10.3,21.1], P=2.0×10^-10). A polygenic score based on results from the Psychiatric Genomics Consortium (PGC) GWAS project was modestly associated with MDD in the testing set (P=6.4x10^-8), accounting for 0.79% of the variability in disease risk. However, when then CONVERGE sample was randomly divided into training and testing sets, and SNP effects estimated by the best linear unbiased prediction (BLUP) method in the training set, the resultant aggregate score was modestly associated with MDD in the testing set (P=3.0x10^-6), accounting for 0.79%; of the variability in disease risk.

Discussion: We have conducted a large GWAS of MDD in an ethnically homogenous sample using SNP data imputed from low-pass sequencing data. In our preliminary analysis, no single variant demonstrated genome-wide significant association with MDD. Aggregate risk scores based on GWAS in European and Chinese populations were found to be of differential predictive value. These observations support a complex etiology for MDD, and possible population differences in predisposing genetic factors.
912W

Calcium, magnesium, potassium and sodium cations as well as the chloride and phosphorus anions are major dietary minerals which can not be manufactured by the human body. They play important roles in various biological functions and are essential to sustain life and maintain optimal health, and thus are commonly measured in blood serum to monitor a range of health conditions. The mineral intake of children and adolescents are especially important due to their rapid growth. In this study, we performed a genome-wide association study to explore the contribution of potential common genetic variations to serum concentrations of the six major dietary minerals in Children as well as the ratio of concentrations between minerals. A total of 10,308 children, 5602 Caucasians and 4706 African-Americans were included in the analysis. All samples were genotyped on the Illumina HH550 or 610 arrays. Maximum serum concentration values of the six minerals, and the mineral ratios, were tested for association by linear regression as implemented in plink. In the African-American pediatric cohort, we report association of the serum calcium concentration with a missense variant in the calcium-sensing receptor (CASR) gene on 3q13 (rs1801725 5.6x10E-3. The SNP was previously reported to correlate with the level of serum calcium in both Caucasian and Indian-Asian populations, but no previous studies show a calcium concentration association in the African-American population or pediatric populations. The association between SNP rs1801725 and calcium concentration was also confirmed in our Caucasian pediatric cohort with a p-value of 1.03x10E-4; combined P-val 2.64x10E-6. Our results extended the understanding of CASR mediated calcium regulation in African-American and Caucasian pediatric populations. In the Caucasian pediatric population, our results also confirmed another previously reported serum magnesium concentration associated SNP rs4072037 with a p-value of 1.74x10E-5 located in gene MUC1 on 1q21. We also investigated the ratio of Calcium/Magnesium, Sodium/Potassium and Chloride/Phosphorus concentrations in blood serum. We report significant association of Calcium/Magnesium ratio with variants located that map to a locus on 10q26 that contains the Fibroblast growth factor receptor 2 gene (FGFR2). Results of the GWAS for the six minerals and the mineral ratios will be presented at the meeting.

913T
GWAS meta-analysis identified a novel locus associated with corneal curvature in Asian populations. P. Chen1, C.Y. Cheng2, T. Aung1,2,3, C.C. Khor1,5, Y.T. Wong4,5, E.S. Tan2,6, E. Tham2,7, Y. Teo1,2,7, A. Cheung8,9. 1) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 2) Singapore Eye Research Institute, Singapore National Eye Center, Singapore; 3) Department of Ophthalmology, National University of Singapore, Singapore; 4) Department of Paediatric National University Health Systems, National University of Singapore, Singapore; 5) Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 6) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore; 7) Duke-National University of Singapore Graduate Medical School, Singapore; 8) Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School, Singapore; 9) NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore; 10) Life Sciences Institute, National University of Singapore, Singapore; 11) Department of Statistics and Applied Probability, National University of Singapore, Singapore.

Corneal curvature (CC) is an ophthalmic biometric that has connection to various eye diseases. CC is highly heritable but there are only two loci discovered currently, FRAP1 and PDGFRA. Both of the two genes have been reported in Asian populations for the first time in an Asian genome-wide association (GWAS) meta-analysis. Although PDGFRA has been well established to be associated with CC, FRAP1 has not been replicated in other populations. Meanwhile in Asian populations, the proportion of variance which is explained by these two genes ranges from 1.8% to 11.1%. We conducted a genome-wide association study meta-analysis of CC to provide further evidence for the association of CC with FRAP1 in additional populations, and discover more genes that are still hiding behind the scene. In total, we recruited 5326 Chinese, 2138 Malay and 2124 Indian individuals in Singapore from 7 GWAS. This study adopted a two-stage meta-analysis scheme. The discovery stage studies (SP2, SiMES, SINDI and SCORM) have been recycled by the previous meta-analysis. Two novel studies, SCES and STARS comprising 2473 Chinese individuals were employed in the replication stage. The radius of CC was defined as the average reading of right and left eye. Within each study, the radius was normalized by an inverse-normal transformation. Imputation of the SNP array genotypes was done using 1000 Genomes project phase 1 (March 2012, GRCh37) cosmo-politan haplotypes as reference panel. Effect of the allele dosage was evaluated in a linear regression model which adjusted for age and gender. Prin-ciple components were also adjusted for Malay and Indians to take care of the population stratification. The meta-analysis was done using a fixed effect inverse-variance weighting scheme. We successfully replicated the signals at FRAP1 and PDGFRA gene region. Notably, the association of FRAP1 was again replicated in our replication studies. Besides the two known genes, we found that CMPK1 gene was associated with CC for the first time (array SNP meta-analysis P value = 2.81E-9), whereas the imputed genotypes brought us even more power with more significant signal (imputed top SNP meta-analysis P value = 3.27E-11). Our meta-analysis replicated the two known genes and identified a novel locus to be associated with CC. We provided further evidence that FRAP1 is associated with CC in Asian populations. However, whether it is an Asian-specific CC susceptibility gene remains to be assessed in other populations.
914F
Uncovering loci associated with urinary incontinence in African and Hispanic American women. C. Chen,1 A. Rajkovic,2 A. Park,3 G. Heiss4, S. Hendrix4, N. Franceschini4. 1) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) department of obstetrics, gynecology and reproductive science, University of Pittsburgh, Pittsburgh, PA; 3) Department of Obstetrics & Gynecology, Medstar Washington Hospital Center, Georgetown University School of medicine, Washington, DC; 4) Department of Epidemiology, Gillings School of Public Health, University of North Carolina, Chapel Hill, NC; 5) Department of Surgical Services, Michigan State University College of Osteopathic Medicine, East lansing, MI.
Urinary incontinence is an under-recognized health problem that negatively impacts women’s quality of life. To date, genes contributing to urinary incontinence susceptibility have not been identified. We present the first genome-wide association study aimed to identify single nucleotide polymorphisms associated with overall urinary incontinence and two subtypes of urinary incontinence (stress/urge) in 11,526 US minority women from Women’s health Initiative SHARe cohort. These women included 8,088 African American and 3,438 Hispanic American women. We report that the prevalence of overall urinary incontinence in these women aged 50-79 years was 55%, with 32% of women experiencing urge urinary incontinence, 29% of women having stress urinary incontinence and 8% of women had both. More African women reported having urge urinary incontinence while more Hispanic women reported having stress urinary incontinence. We show that rs2086297, located in the intron of PRCP, was significantly associated with stress UI (p-value<4.4×10^-8) in trans-ethnic analysis. Each copy of the rs2086297, located in the intron of PRCP, was significantly associated with having stress urinary incontinence and 8% of women had both. More African women were associated with urinary incontinence while more Hispanic women were associated with stress UI. The strongest evidence was observed at chromosome 21q22 (β=0.08; t=2.571) were performed. Using random effect models, the pooled estimates of the four cohorts confirmed the association at the 21q22 locus (β [the increase in nuclear cataract grade per risk allele] = 0.11, P = 5.2 ×10^-9), reaching genome-wide significance. The results were similar after further adjustment for covariates. Replication tests in two independent Chinese cohorts (n = 2,571) were performed. Using random effect models, the pooled estimates of the four cohorts confirmed the association at the 21q22 locus (β = 0.08, P = 2.0 ×10^-9). To our knowledge, this is among the first GWAS to investigate age-related nuclear cataract. The results yield insights at the gene level for the pathogenesis of age-related cataract.

915W
Genome-wide association study of thyrotoxic periodic paralysis. K.C. Chen1, P.L. Chen1,2,3,4, T.C. Chang2,3,4, 1) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan; 5) Department of Internal Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan.
Thyrotoxic periodic paralysis (TPP) is a unique disease characterized by episodic attacks of muscle weakness and hypokalemia in certain thyrotoxic individuals. TPP shows male predominance and huge inter-population prevalence difference; up to 10% of thyrotoxic males do. In 2010, Ryan et al. (Cell 140(1): 88-98) reported that mutations of the KCNJ18 gene could be found in 33% of their patients (mainly of European descent), but not in controls. However, in 2012, Cheung et al. (Nat Genet 44(9): 1026-1029) published a genome-wide association study (GWAS) in ethnic southern Chinese and identify a single susceptibility locus at 17q24.3 near KCNJ2, with estimated odds ratio of 3.3. It is still an unsolved issue whether these two loci are unique TPP-related loci, and whether there are more susceptibility loci to be identified. To address these issues, we conducted a GWAS using 41 TPP patients as cases and 725 thyrotoxic, non-TPP individuals as controls, all of them are ethnic Chinese in Taiwan. The genotyping platform was Axiom Genome-Wide CHB Array Plate, which contains 842,832 common (MAF>5%) SNPs in Han Chinese genome. We replicated the 17q24.3 association signals reported by Cheung et al., with several SNPs showed p values smaller than 5 ×10^-8 in the trend test and allelic test. Fine mapping of the novel loci is currently underway. On the other hand, we did not find association signals at or near KCNJ18. We are now performing Sanger sequencing of KCNJ18 to search for possible multiple rare variants. Our results demonstrate that 17q24.3 is the genuine TPP susceptibility locus, at least in ethnic Chinese Han population. We also report several novel susceptibility loci awaiting replication.

916T
Common variations at chromosome 21q22 influence the risk of age-related nuclear cataract in Asians: the Singapore Epidemiology of Eye Diseases (SEED) Study. C.Y. Cheng1,2,9, J.J. Wang1,2,9, P. Cheng1, X. Li1,4, X. Wang2, A.G. Tan1, J.J. Wang2,6, P. Mitchell8, J.B. Jonas1, S.M. Saw3, C.C. Khor1, E.S. Tai1,2, T. Aung1,3, Y.Y. Teo7,4,11,12, T.Y. Wong1,3,10.
1) Department of Ophthalmology, National University of Singapore, Singapore; 2) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 3) Singapore Eye Research Institute, Singapore; 4) Department of Statistics and Applied Probability, National University of Singapore, Singapore; 5) Centre for Vision Research, Department of Ophthalmology, University of Sydney, New South Wales, Australia; 6) Center for Eye Research Australia, University of Melbourne, Melbourne, Australia; 7) Department of Ophthalmology, Medical Faculty Mannheim, Ruprecht-Karls-University Heidelberg, Mannheim, Germany; 8) Human Genetics, Genome Institute of Singapore, Singapore; 9) Department of Medicine, National University of Singapore, Singapore; 10) Singapore National Eye Centre, Singapore; 11) Centre for Molecular Epidemiology, National University of Singapore, Singapore; 12) Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore.
Age-related cataract is the leading cause of visual impairment and blindness in the world. Although several mutations were identified as causative for congenital and juvenile cataracts, little is known about the genetic variants that influence the susceptibility of age-related cataract. We conducted a meta-analysis of genome-wide association studies (GWAS) on 2,369 Malays and 2,200 Indians aged 40 to 80 years enrolled in the Singaporean Epidemiology of Eye Diseases (SEED) Study. Participants underwent a comprehensive eye examination including slit-lamp lens photography and provided a blood sample for genotyping. Lens photographs were graded using the Wisconsin Cataract Grading System, with the grades ranging from 0.1 (the lowest grade) to 5.0 (the highest). Genotyping was performed with Illumina HumanHap610-Quad chips. We performed SNP imputation using the genoty- pe data together with the reference panels in the 1000 Genomes project (March 2012, GRCh37). The association between SNPs and the severity (i.e., grade) of nuclear cataract was assessed using linear regression analysis with adjustment for age, sex and genetic principal components. In the multi-ethnic analysis of genome-wide association studies (GWAS) on 2,369 Malays and 2,200 Indians aged 40 to 80 years enrolled in the Singaporean Epidemiology of Eye Diseases (SEED) Study. Participants underwent a comprehensive eye examination including slit-lamp lens photography and provided a blood sample for genotyping. Lens photographs were graded using the Wisconsin Cataract Grading System, with the grades ranging from 0.1 (the lowest grade) to 5.0 (the highest). Genotyping was performed with Illumina HumanHap610-Quad chips. We performed SNP imputation using the genoty- pe data together with the reference panels in the 1000 Genomes project (March 2012, GRCh37). The association between SNPs and the severity (i.e., grade) of nuclear cataract was assessed using linear regression analysis with adjustment for age, sex and genetic principal components. In the multi-ethnic analysis on 2,369 Malays and 2,200 Indians aged 40 to 80 years enrolled in the Singaporean Epidemiology of Eye Diseases (SEED) Study. We replicated the association at the 21q22 locus (β [the increase in nuclear cataract grade per risk allele] = 0.11, P = 5.2 ×10^-9), reaching genome-wide significance. The results were similar after further adjustment for covariates. Replication tests in two independent Chinese cohorts (n = 2,571) were performed. Using random effect models, the pooled estimates of the four cohorts confirmed the association at the 21q22 locus (β = 0.08, P = 2.0 ×10^-9). To our knowledge, this is among the first GWAS to investigate age-related nuclear cataract. The results yield insights at the gene level for the pathogenesis of age-related cataract.
917F Copy Number Variations are associated with Bone Mineral Density: A large-scale genome-wide analysis in the Framingham Study. W. Chou1, K. Nandakumar1, D. Karasaki1, C. Liu4, L. Cupples1, D. Kiel1, Y. Hsu1, 1. Institute for Aging Research, Hebrew SeniorLife, Harvard Medical School, Boston, MA; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Bone mineral density (BMD) is a complex phenotype with high heritability. Our previous genome-wide SNP association study identified more than 56 BMD loci in Caucasian populations, explaining <6% of BMD variation. Copy number variation (CNV), a type of genomic structural variation, accounts for >20% of the variance of the human genomic structure between individuals. CNVs may explain the missing heritability of BMD. Thus, we conducted a genome-wide CNV association analysis with BMD in Framingham Study participants. This study included 7,451 adult Caucasians (4126 men and 3325 women) with mean age of 55 years. BMD at lumbar spine (LS) and femur neck (FN) was measured by dual energy X-ray absorptiometry. CNVs were estimated using Affymetrix 500K genotyping array and PennCNV package. A CNV was defined as a DNA segment longer than 1 kb and composed of at least three consecutive genotyping probes. Under an additive genetic model, we employed a linear mixed effects model to account for family relatedness, and also adjusted for age, sex, estrogen usage, menopause status, cohorts within the Framingham Study and principal components for population stratification. To correct for multiple-testing, a genome-wide significant cutoff (p < 4 × 10−5) was defined by a false-discovery rate as less than one false-positive result among the genome-wide significant findings. Up to 6,398 individuals had valid CNV calls and information on covariates. Two CNVs (at chrom 2q14.1 and 6q25.3) and four (at chrom 5p22.1, 15q22.31, 11q22.1 and 1q21.2) were significantly associated with BS BMD and FN BMD, respectively. The length of the CNVs ranged from 133 kb to 1,025 kb. The most significantly associated CNV with p = 1.9 × 10−20 was located at chromosome 15q22.31 (SMAD6). The Smad6 protein inhibits signaling of bone morphogenetic proteins. Among 6,398 individuals included in the association analysis of the CNV, 5 had this CNV. Individuals with deleted CNVs had lower FN BMD. In addition, we identified three genome-wide suggestive CNVs associated with BMD at p < 2.7 × 10−4. In summary, we identified six genome-wide significant CNVs and three genome-wide suggestive CNVs associated with FN BMD or BS BMD in the Framingham Study. To replicate the CNV-BMD associations, we are performing analyses in independent samples. The newly identified CNVs in this study may provide additional information on genetic determinants of BMD and may explain some of the missing heritability.

917T Exploring the causes of heterogeneity in meta-analysis of genome-wide association studies. H. Deng1,2, Y. Pei1,2, Z. Zhagn1,2, 1) Biostatistics and Bioinformatics, Tulane Univ, New Orleans, LA; 2) Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

Meta-analysis of Genome wide association (GWA) data has been widely used to detect genetic variants responsible for complex diseases. Between-study heterogeneity is an important issue when conducting meta-analysis. Exploring the extent and possible sources for heterogeneity between studies is important to the result interpretation. To exemplify these issues, we used data from seven GWA studies on obesity. We first investigated the extent and distribution of the heterogeneity across the whole genome. Among the 4,325,550 SNPs tested, the I2 inconsistency metric was different from 0 for 43.8% of the total SNPs; inconsistency was moderate to very large (I2>25%) for 18.4% of the total SNP. Significant heterogeneity (p<0.1) existed for 11.0% of the total SNPs. Heterogeneity was enriched in SNPs with small p-values. Among SNPs with p-value less than 1.0×10−5, 190 (41.4%) SNPs showed moderate or larger heterogeneity. Meta-regression analyses were performed for the 190 SNPs to explore potential causes of heterogeneity. Five moderators, which included average age, ethnicity, gender composition, study base and imputation accuracy, were constructed to test their effects on heterogeneity. Among the tested factors, ethnicity, average age and imputation accuracy had significant effects on the heterogeneity. After adjusting for all the five moderators, almost all of the significant heterogeneity (96.2%) was removed. Our results have significant implications for the study design and results interpretation of GWA meta-analyses.

920F Cross-disease analysis using Immunochip reveals four new loci for celiac disease and rheumatoid arthritis. J. Gutierrez-Achury1, G. Tryfka1,2, S. Raychaudhuri1,2, J. Greenberg1,2, D. Diogo3, R. McManus3, R.M. Plange4,5, C. Wijmenga3,6, K. Nandakumar7, D. Karasaki9, A. Custovic10, Y. Hsu11, J.A. Curtin12. 1) Department of Genetics, University Medical Centre Groningen - University of Groningen, Groningen, Netherlands; 2) Division of Genetics and Rheumatology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA; 3) Medical and Population Genetics Program, Broad Institute, Cambridge, MA, USA; 4) Institute of Inflammation and Repair, University of Manchester, Manchester, UK; 5) New York University School of Medicine, New York, NY; 6) Department of Clinical Medicine and Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland.

Background: Autoimmune diseases are complex and heterogeneous but are known to share a subset of loci and pathways. Celiac disease (CeD) and rheumatoid arthritis (RA) are two examples in which GWAS and the immunochip have identified strong associations to each trait and substantial overlap between loci. The Immunochip array was specifically designed to refine GWAS results. We aimed to jointly analyze the largest CeD and RA case-control datasets available to date and genotyped using the Immunochip to identify new unique or shared loci and to fine-map the known regions. Methods: We meta-analyzed 6 European CeD cohorts (12,400 cases, 14,257 controls) and 7 European and North American RA cohorts (13,877 cases, 18,120 controls). We determined common SNPs between the datasets, aligned them in the same strand, removed duplicates and related individuals, and performed a principal component analysis (PCA) per cohort. We performed logistic regression by adjusting for first five PCAs per cohort. We then performed a principal component analysis (PCA) per cohort. We performed logistic regression by adjusting for first five PCAs per cohort. We then defined meta-analyses using a fixed-effect model with inverse variance weighting. Results: From the joint meta-analysis of all 13 cohorts, 28 loci reached the threshold p-value of association, 1×10−8, and 17 more reached a suggestive p-value of 1×10−6. Of those that reached genome-wide significance, we report the new locus NCF2 (ORmax=1.25, p-valmax=1.24×10−14) represented by a probably damaging missense mutation with a polyphen-2 score of 0.98. Three other loci (AFF3, PTPN22 and CXCR5) has been previously reported in one disease but were now identified in the second phenotype and in the combined analysis. Using co-expression analysis based on approximately 80,000 microarrays, it was possible to determine that the new gene NCF2 is involved in the defense response to Gram-positive bacteria, which points to its importance in the innate immune response and microbiota regulation. These aspects should be further analyzed by functional studies.
921W

Asthma and allergic disease are common, chronic conditions with substantial public health burdens. We carried out a genome-wide association study of self-reported asthma in the 23andMe participant cohort, including more than 17,000 cases and 87,000 controls with European ancestry, imputed against 1000 Genomes reference haplotypes. In addition to replicating established asthma GWAS findings, we see genome-wide-significant associations with several loci that have been associated with other forms of atopic disease, including rs17635535 in CLEC16A, rs1254017 near ZBTB10, and rs3001426 near STAT6. Using conditional logistic regression, we see evidence for multiple independent associations at the IL1R1 locus on chromosome 2, the TSLP locus on chromosome 5, and the IL33 locus on chromosome 9.

We find novel associations with rs2070902 in FCER1G, encoding the gamma subunit of the high affinity IgE receptor; with rs4707690 in BACH2, or basic leucine zipper transcription factor 2; and with rs62192043 near D2HGDH, or mitochondrial D-2-hydroxyglutamate dehydrogenase, and GAL3ST2, or galactos-3-O-sulfotransferase. Variation in BACH2 has previously been associated with a variety of autoimmune diseases. GAL3ST2 is expressed in goblet cells and is involved in production of sulfated mucins. We also find evidence of associations in or near several genes associated with allergies in the 23andMe cohort, including ID2, LPP, TLRI, and PTGER4.

Most genetic associations with asthma and allergy are shared, with similar effect sizes, and identifying risk factors that distinguish between different allergic disease types seems more challenging. We further explore this issue with analyses of the susceptibility loci stratified by disease subgroups, determined by age of onset, triggers, and severity, and with patterns of allergy symptoms.

922T
A genome-wide association study highlights multiple variants associated with Epstein-Barr virus load in the 1000 Genomes and HapMap lymphoblastoid cell lines. C.J. Houldcroft1, J.Z. Liu1, A. Gall1, D. Framp1, C.A. Anderson2, P. Kollias2,1, Welcombe Trust Sanger Institute, Cambridge, Cambs, United Kingdom; 2) UCL/MRC Centre for Medical Molecular Virology, Department of Infection, University College London, Cleveland Street, London W1T 4JF, UK.

Epstein-Barr virus (EBV) infects 95-98% of adults world-wide, persists for life and causes benign infectious mononucleosis in ~50% of young adults upon primary infection. EBV is also linked to more severe conditions such as Burkitt’s lymphoma, nasopharyngeal carcinoma and post-transplant lymphoproliferative disorder. Persistent life-long infection can be modelled using lymphoblastoid cell lines (LCLs), human B cells immortalised with EBV. It is unknown to what extent host genetic variants influence EBV load within LCLs and whether variation in EBV load affects gene expression studies of these cell lines.

We measured relative EBV genome load in LCLs from 1000 Genomes and HapMap projects and performed a heritability analysis and genome-wide association study (GWAS) on 899 of these individuals. This combined sample has ~150 SNPs associated with EBV load genome copy number, and is enriched for cSNPs affecting miRNA expression of FCGR2A. We also established the genetic association between IL12B and EBV load (rs50617332, OR= 1.54, P= 2.18× 10−8). An association with an additional locus on chromosome 21q22 downstream of PMSG1 did not pass the threshold for genome-wide significance (P=7.39×10−8).

Conclusion: We established multiple genetic susceptibility loci for Takayasu’s arteritis with a genome-wide level of significance including two independent susceptibility loci in the HLA region, and disease susceptibility loci in FCGR2A/FCGR3A and IL12B.

923F
Identification of multiple genetic susceptibility loci in Takayasu’s arteritis. T. Hughes1, G. Saruhan-Direskeneli2, P. Cot3, J.M. Guthridge3, J.A. James4, P. Merkely5, H. Direskeneli2, A.H. Sawalha6,1, on behalf of Vasculitis Clinical Research Consortium,* on behalf of Turkish Takayasu Study Group, 1) Division of Rheumatology, University of Michigan, Ann Arbor, MI, USA; 2) Department of Physiology, Istanbul University, Istanbul Faculty of Medicine, Istanbul, Turkey; 3) DCRI Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA; 4) Division of Rheumatology, University of Pennsylvania, Philadelphia, PA, USA; 5) Department of Rheumatology, Marmara University, Faculty of Medicine, Istanbul, Turkey.

Background: Takayasu’s arteritis is a rare inflammatory disease of large arteries. The etiology of Takayasu’s arteritis remains poorly understood, but genetic contribution to the disease pathogenesis is supported by the strong genetic association with HLAB.5. Genetic studies in Takaya-

su’s arteritis have not been previously performed. Methods: We genotyped ~200,000 genetic variants in two ethnically divergent Takayasu’s arteritis cohorts from Turkey (339 patients and 516 controls) and North America (112 patients and 599 controls) using a custom designed genotyping platform (ImmunoChip). Additional genetic variants and the classical HLA alleles were imputed and analyzed. Results: We identified and confirmed two independent susceptibility loci within the HLA region (rs2<0.2): HLA-B/MICA (rs12524487, OR= 3.29, P= 5.57× 10−8), and HLADQB1/HLA-DRB1 (rs113452171, OR= 2.34, P= 3.74× 10−8, and rs189754752, OR=4.22× 10−10). In addition, we identified and confirmed a novel genetic association between Takayasu’s arteritis and the FCGR2A/FCGR3A loci on chromosome 1 (rs10919543, OR= 1.81, P= 5.89× 10−11). The risk allele in this locus results in increased mRNA expression of FCGR2A. We also established the genetic association between IL12B and Takayasu’s arteritis (rs56167332, OR= 1.54, P= 2.18× 10−8). An association with an additional locus on chromosome 21q22 downstream of PMSG1 did not pass the threshold for genome-wide significance (P=7.39×10−8).

Conclusion: We established multiple genetic susceptibility loci for Takayasu’s arteritis with a genome-wide level of significance including two independent susceptibility loci in the HLA region, and disease susceptibility loci in FCGR2A/FCGR3A and IL12B.

924W
Genome-wide association confirms TCF4 as a major locus for Fuchs Endothelial Corneal Dystrophy and identifies novel loci. S.K. Iyengar1,2, R.P. Igo, Jr.1, A. Gall1, D. Framp1, C.A. Anderson2,1, 1) Dept Epid/Biostat, Case Western Reserve Univ, Cleveland, OH; 2) Dept, Ophthalmology, UCSD, San Diego, CA; 3) Department of Biostatistics & Bioinformatics, Duke University Medical Center Center for Human Genetics, Duke University Medical Center, Durham, NC; 4) Department of Ophthalmology and Visual Sciences, Case Western Reserve Univ, Cleveland, OH; 5) Duke Eye Center, Duke University Medical Center Dur-

ham, NC.

Background. Fuchs Endothelial Corneal Dystrophy (FECD) is a progressive, age-dependent disease of the corneal endothelium, associated with decreased vision. Women are more commonly affected than men. Histological hallmarks of the disease include decreased endothelial cells, thickened Descemet’s membrane (DM), excessive accumulation of pachymeniscal matrix, and the formation of excrescences on DM called corneal guttae. FECD is one of the most common indications for corneal transplants in developed countries, with a prevalence of approximately 4%, and heritability in excess of 40-50%. Standard methods. We organized a multi-center consortium, the FECD Consortium, which has conducted the largest genome-wide association study to date (Baratz et al. (2010) NEJM 363, 1016). Besides TCF4, five other loci on chromosomes 1q and 11p also reached genome-wide significance with minor-allele odds ratios ranging from 0.75 to 1.43; the latter are significant in the Male subset (P< 0.01, min P=2.6×10−5). Further analyses of histopathology validated cases confirmed all loci. TCF4 shows genome-wide significance in the Male subset (P< 0.001), and reaches genome-wide significance in the Female subset (P< 0.001 with min P= 2.8×10−8), but two loci show promising results in the Male subset (P< 0.01, min P=2.6×10−5). Markers at all six loci yielded an AUC of 0.77, but the majority of the predictive value came from TCF4 (AUC=0.78). Besides TCF4, the FECD Genetics Consortium has conducted the largest genome-wide association study to date on FECD, leading to confirmation of the role of TCF4 on FECD. Five novel loci on chromosomes 1 and 11 were identified. Further analyses of these genes should help elucidate different pathways in FECD pathogenesis.
Sub-phenotype Mapping in Systemic Lupus Erythematosus Identifies Multiple Novel Loci Associated with Circulating Interferon Alpha. S. Kariuki1, Y. Ghodke-Puranik1, J. Dorschner1, B. Chrabot2, J. Kelly1, B. Tsao2, R. Kimberly1, M. Alarcón-Riquelme3,6, C. Jacob2, L. Criswell2, K. Sivils3, C. Langefeld1, J. Harley3, A. Skol1, T. Niewold2. 1) University of Chicago, Chicago, IL; 2) Mayo Clinic, Rochester, MN; 3) Oklahoma Medical Research Foundation, Oklahoma City, OK; 4) University of California, Los Angeles, CA; 5) University of Alabama, Birmingham, AL; 6) Pfizer - Universidad de Granada - Junta de Andalucia Centre for Genomics and Oncological Research (GENYO), Granada, Spain; 7) University of Southern California, Los Angeles, CA; 8) University of California, San Francisco, CA; 9)Wake Forest University, Winston Salem, NC; 10) Cincinnati Children's Hospital Medical Research Center, Cincinnati, OH.

Systemic Lupus Erythematosus (SLE) is a phenotypically heterogeneous complex disease. Our previous work has documented significant genetic heterogeneity, with some well-validated risk factors demonstrating strong sub-group effects. Approximately 50% of patients have high circulating levels of interferon alpha (IFN-α), and many lines of investigation suggest IFN-α as a heritable and primary causal factor in human SLE. This study aims to genetically map the serum IFN-α trait in SLE patients, allowing for novel genetic discovery in this heterogeneous disease. GWAS data were obtained from 450 European ancestry SLE cases who were genotyped as part of the Systemic Lupus Erythematosus Genetics (SLEGEN) study. Genotypes were generated on the Illumina Infinium HumanHap300 genotyping platform, and principal component analysis was used to correct for population stratification. Sera were obtained from each of these subjects, and IFN-α activity was measured using a sensitive and specific reporter cell assay. Associations between genotype-wide SNP markers and serum IFN-α were detected using logistic regression conditioned on the principal components to control for structure. IFN-α activity was studied as a categorical trait. Patients with IFN-α levels two standard deviations above the mean of healthy controls were designated as high IFN-α, and the remainder as low IFN-α. Top novel associated loci in the GWAS screen include multiple SNPs in the C7orf57, PRKG1, ANKRD44, and NPN loci. Interestingly, three of the five top SNPs are missense SNPs. Strong association signals were also detected in chromosome 9, which is currently underutilized. This is largely due to exclusion of X chromosome from analyses even though these regions are assayed on many current microarray platforms. Employing genotype and phenotype data from 6 European ancestry cohorts, we imputed the genotypes of 9,511 controls and 8,497 cases to test for genome wide associations for Parkinson's disease on the X chromosome. While a plethora of GWA data is available on the X chromosome, it is currently underutilized. This is largely due to exclusion of X chromosome variants from analyses even though these regions are assayed on many current microarray platforms. Employing genotype and phenotype data from 6 European ancestry cohorts, we imputed the genotypes of 9,511 controls and 8,497 cases to test for genome wide associations for Parkinson's disease on the X chromosome. While a plethora of GWA data is available on the X chromosome, it is currently underutilized. This is largely due to exclusion of X chromosome variants from analyses even though these regions are assayed on many current microarray platforms. Employing genotype and phenotype data from 6 European ancestry cohorts, we imputed the genotypes of 9,511 controls and 8,497 cases to test for genome wide associations for Parkinson's disease on the X chromosome.
928T
Genetic variants associated with total lung capacity in chronic obstructive pulmonary disease. J. Lee1,2, G.M. Hunninghake3, M.N. McDonald1, M.H. Choi1,2, M. Hardin1,2, E.S. Wain1,2, P.J. Castaldi1, A. Gulsvik1, P. Bakke1, E.K. Silverman1,2, C.P. Hersh1,3, The COPDGene and ECLIPSE investigators. 1) Channing Division of Network Medicine, Brigham and Women’s Hospital, Boston, MA; 2) Department of Internal Medicine, School of Medicine, University of California, San Francisco, San Francisco, CA; 3) University of Pittsburgh School of Medicine, Ewha Womans University, Seoul, Korea; 3) Pulmonary and Critical Care Division, Brigham and Women’s Hospital, Boston, MA; 4) Haukeland University Hospital and Institute of Medicine, University of Bergen, Bergen, Norway.

Background: Chronic obstructive pulmonary disease (COPD) is characterized by expiratory flow limitation, resulting in air trapping and lung hyperinflation. Hyperinflation is closely related to exercise tolerance and quality of life in COPD patients. It could also be used as a distinguishing feature to explain COPD heterogeneity. The aim of this study was to identify genetic variants associated with hyperinflation in COPD. Methods: We performed genome-wide association studies in three cohorts: COPDGene; the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE); and GenKOLS (Bergen, Bergen, Norway). All subjects were Caucasians and had at least moderate COPD, defined by a ratio of forced expiratory volume in 1 second to forced vital capacity (FEV1/FVC) <0.7 and FEV1 <80% predicted in post-bronchodilator spirometry. Total lung capacity (TLC) was used as a marker of hyperinflation and calculated by using volumetric computed tomography scans of the chest. Genotyping in each cohort was performed on the Illumina platforms, with additional markers imputed using 1000 Genomes CEU or EUR reference panels. To find genetic variants associated with TLC, we used linear regression models adjusting for age, sex, pack-years of smoking, height, and principal components of genetic ancestry in PLINK version 1.07, with meta-analysis of the results using METAL. Results: The most significant loci were on 17p13.2 (β= -2.58; p=2.51×10^-7), 3q24 (β= -1.22; p=9.64×10^-7), and 19p21.3 (β= -1.22; p=9.64×10^-7) in the COPDGene cohort; 17p13.2 (β= -1.93; p=3.34×10^-7) and 19q23.3 (β= -1.22; p=9.64×10^-7) in the ECLIPSE cohort; and 8p11.21 (β= -2.58; p=2.51×10^-11), 3q24 (β= -1.87; p=4.88×10^-7) and 9q34.13 (β= -1.97; p=5.05×10^-7) in GenKOLS cohort. In a meta-analysis of these three cohorts, the most significant loci were on 17p13.2 (β= -1.71; p=1.82×10^-7) and 19p13.12 (β= -0.13; p=2.36×10^-7). Conclusions: Although no single genetic variant reached genome-wide significance, there were several suggestive associations with TLC in three COPD cohorts. Expanding to another cohort could improve statistical power to elucidate associations between genetic variants and hyperinflation-related phenotypes. Funding: NIH R01 HL089856 and R01 HL089897.

929F
The genome-wide association analysis and meta-analysis of sciatica in two Finnish populations. S. Lemmelä1, S. Solovieva1, R. Shiri1, J. Seppälä2,3, M. Hellöövaara4, M. Kahönén2,3, M. Juonala2,5, J. Vilkan5,6, O. Ratakari5,6, T. Lehtimäki1,7, E. Vikrant-Juntura1, K. Husgafvel-Pursiainen1,1,1, 1) Finnish Institute of Occupational Health, Helsinki, Finland; 2) Tampere University Hospital, Tampere, Finland; 3) University of Tampere, Tampere, Finland; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) University of Turku, Turku, Finland; 6) Turku University Hospital, Turku, Finland; 7) Fimlab Laboratories, Tampere, Finland.

Sciatic pain is one of the leading causes of disability in the working age population. A strong genetic component in family and twin studies as well as a multifactorial etiology as documented in epidemiological studies have been shown for sciatica. A genome-wide association study was conducted in two large population-based Finnish cohorts informative for sciatica. In our current study, data containing more than 500 000 genotyped SNPs and several million imputed SNPs as based on 1000 Genomes imputation reference constructed by the IMPUTE program were analysed using PLINK and SNPTEST, respectively. Results were adjusted for age and gender. Finally, meta-analysis using result data of both cohorts was conducted using GWAMA. In summary, our preliminary results revealed novel candidate regions for sciatica. These results, with several highly interesting regions warranting further investigation, will be discussed. The study received financial support from the Academy of Finland (project no 129364, MSDs@Lifecourse, The SALVEcourse).

930W
Replication of GWAS results for dental caries in the permanent dentition. D. Lewis1, J. Shaffer1, E. Feingold1,2, M. Cooper3, R. Weyant1, D. McNeil3, R. Crouse1, S. Reis1, A. Vieira4, M. Vanyukov1, M. Marazita1,3,6, 1) Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 2) Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 3) Center for Craniofacial & Dental Genetics, Department of Oral Biology, School of Dental Medicine, Pittsburgh, PA; 4) School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 5) Dental Practice and Rural Health, West Virginia University, Morgantown, WV; 6) Clinical and Translational Science, School of Medicine, University of Pittsburgh, Pittsburgh, PA; 7) University of Pittsburgh School of Pharmacy, Pittsburgh PA; 8) Health Policy Institute, University of Pittsburgh, PA.

There is increasing evidence that variation in the risk for dental caries (i.e., tooth decay) includes a genetic component. We previously published several GWAS studies for dental caries in the permanent dentition (adult teeth), primarily based on the dataset from the GENEVA dental caries study. Those publications considered several different phenotypes, including overall number of carious teeth in the mouth(DMFT)and numbers of carious lesions in specific types of tooth surfaces. Interesting nominated genes included BMP4, BCOR, LYZL2, and ABCG2. BMP4 is known to initiate and regulate the repair of carious tissue. Mutations in BCOR cause oculoauriculovertebral syndrome. In the current work, we did follow-up genotyping in suggestive regions in four study populations totaling approximately 2000 individuals in order to replicate and fine-map our original results.

Grants
R01-DE014899
U01-DE198903.

931T
Negative-regulation-of-apoptotic-process identified to be associated with appendicular lean mass through meta-analysis of pathway-based genome-wide association analysis. J. Li1, C. Xu2, YF. Pei1,2, Q. Tian1, H-W. Deng1,2. 1) Center for Bioinformatics and Genomics, Biostatistics and Bioinformatics, Tulane University, New Orleans, LA, 70112 USA; 2) Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

Background: Sarcopenia, characterized by degenerative loss of skeletal muscle mass and strength with aging, is a major public health problem. It is a heritable disease with the heritability of muscle mass and strength being estimated as high as ~60%. Body lean mass (LM) and appendicular lean mass (aLM, the total muscle mass in limbs) are often used as measures for sarcopenia. Recent studies, especially, genome-wide association studies (GWAS), had linked a few genes to LM or aLM. However, they only account for a small proportion of the total LM/aLM variation. Two potential approaches for better identification of genes, especially those with relatively small effects, related to LM/aLM are meta-analyses and pathway-based analysis. In this study, we had conducted combined meta-analysis and pathway-based association analysis for LM and aLM. Methods: Using seven population samples with over 10,000 subjects and leveraging on the genomic variant information from 1000 Genomes Project, we performed GWAS on the imputed genotypes within each sample. We then conducted pathway-based analyses on the GWAS results through GenGen, before conducting meta-analysis on the pathway-based analysis results. Results: The negative regulation of apoptotic process (NROAP) pathway was identified to be the most significant pathway associated with aLM (raw p-value 1.06×10^-8 for 885 pathways after QC). The pathway had q-value less than 0.05 in the majority of the samples in which aLM measurements were available. Previous studies had indicated that NROAP included genes for negative regulation of muscle cell and mesenchymal cell apoptotic process, potentially affecting the muscle mass changes in the human body. Conclusion: In this study, we identified the potential role of NROAP pathway in human aLM variation and sarcopenia. These results provided new insights into the genetic basis of osteoporosis, and potential targets for further functional analysis.
Cannabis Dependence Is Associated with Genetic Variants in Genes Linked to Schizophrenia.

M. Lohsen1,2,3, R. Sherva1,2, R. Koesterer1, H. Kranzler1, H. Zhao5, L. Almasy5, L. Farrer1,2,3, J. Gelemer2,3,1) Boston University School of Medicine, Department of Medicine, Boston, MA; 2) Boston University School of Medicine, Department of Medicine, Graduate Program in Molecular Medicine, Boston, MA; 3) Boston University, Transforming Training Program in Addiction Sciences, Boston, MA; 4) University of Pennsylvania School of Medicine, Department of Psychiatry and VISHN MIRECC, Philadelphia VAMC, Philadelphia, PA; 5) Yale School of Public Health, New Haven, CT; 6) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 7) Yale University School of Medicine, Departments of Psychiatry, Genetics and Neurobiology, New Haven, CT.

Background: Cannabis dependence (CaD) is a major public health problem, but little is known about the genetic risk factors. A case-control GWAS on a sample of 708 cases and 2364 exposed controls from the Study of Addiction: Genetics and Environment (SAGE) revealed suggestive evidence of association with the ANKFRN1 gene (Agrawal et al., 2011), which were not genome-wide significant (GWS). Our study provides data from a larger sample with a discovery data set and meta-analysis using results from SAGE. Methods: To determine potential links between genetic variants and addiction, users of various substances (alcohol, cocaine and opioids) were recruited in multiple U.S. cities. Many had comorbid substance dependences as well, including CaD. The Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) was administered for psychiatric diagnosis. 988,306 autosomal SNPs (889,659 usable after QC) were genotyped on the IlluminaOmni1-Quad microarray. IMPUTE2 was used to impute genotypes for 10.07 million SNPs for analysis. A GWAS was performed using the summation of the 7 DSM-IV CaD diagnostic criteria in the discovery and replication samples and combined by meta-analysis. Samples consisted of cannabis-exposed controls, cases meeting 3 or more of the 7 DSM-IV diagnostic criteria and others (unexposed or meeting fewer than 3 criteria). African Americans (AAs) discovery: N=3318; 895 cases, 1470 controls, 553 others; SAGE: N=1311; 289 cases, 756 controls, 226 others) and European Americans (EAs) discovery: N=2752; 258 cases, 1589 controls, 705 others) were analyzed separately in each subsample and then combined by meta-analysis. Results and Discussion: We identified several GWS and near GWS associations. The most robust findings were with relatively uncommon variants in the PI4K2B gene for both AAs and EAs (rs313544; pmeta=2.93e-08). The direction of effect was the same for all 4 groups (AAs and EAs, both in discovery and SAGE). This gene was associated with schizophrenia (Houlihan et al., 2009). After NRG1 (Han et al., 2012), this is the second schizophrenia-associated gene we found to be associated with CaD as well. Several other common SNPs approached, but did not reach, GWS, in one or more of the samples. This study presents additional evidence for biological convergence between schizophrenia and substance dependence traits.
935F Replication Study of Age-Related Macular Degeneration Susceptible Gene Using Large Genome-Wide Association Study of Japanese. M. Miyake1,2, K. Yamashiro1, H. Nakashita1,2, I. Nakata1,2, Y. Akagi-Kuras hige1,2, K. Kumagai1, M. Oishi1, A. Oishi1, N. Gotoh1, A. Tsujikawa1, M. Saito2, Y. Kurimoto3, T. Kawaguchi4, CC. Khor2, CY. Cheng2, TY. Wong2, R. Yamada2, F. Matsuda4, N. Yoshimura1, the Nagahama Study Group. 1) Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Center for Genomic Medicine, Inserm U.852, Kyoto University Graduate School of Medicine, Kyoto, Japan; 3) Fukushima Medical School, Fukushima, Japan; 4) Kobe City General Hospital, Kobe, Japan; 5) Genome Institute of Singapore, Singapore, Singapore.

Age-related macular degeneration (AMD) is a major cause of progressive, irreversible visual impairment among elderly population in developed countries. To date, more than 10 genome-wide association studies (GWASs) have been conducted on AMD. In 2013, a large international GWAS (≥ 70,000 subjects) have identified 7 novel AMD susceptible genes (The AMD Gene Consortium, Nature Genetics). However, since this study mainly consists of Caucasians (≥ 94%), this result needs to be replicated in other ethnicity. In the current study, we performed GWAS using large cohort of Japanese to look up the results. Case group consists of 1576 wet AMD patients recruited from all over Japan. These samples were genotyped using Illumina HumanOmni2.5M or OmniExpress. For the control group, we used 3248 healthy Japanese individuals recruited from the Nagahama Prospective Genome Cohort for the Comprehensive Human Bioscience (The Nagahama Study) dataset. After our standard quality control, 558,850 SNPs were included. GWAS was conducted with an adjustment for age, sex, and first 7 principal components. Inflation factor lambda of 1.07 suggested that population stratification was well adjusted. AMRS2 and CFH, that are the established disease susceptible genes of AMD, reached genome-wide significance (P = 4.7x10−7, 1.8x10−3, respectively). We screened ±25kb region of the 7 single nucleotide polymorphisms (SNPs) newly reported by The AMD Gene Consortium. Though p-value of ≤0.05 was observed in ADAMTS9 (lowest P-value = 3.5x10−3, 1.8x10−3, respectively), and SLC16A8 (lowest P-value = 1.8x10−2), 8 additional genetic regions with P values below 10−4. Moreover, there were 38 additional NVR (P = 9.6x10−5) was calculated at each imputed variant over multiple comparison correction. Since no SNP was genotyped within ±50kb around rs3103783 which is located between IER3 and DDR1, we screened ±50kb region for this SNP. This analysis revealed significant association of this region (lowest P-value = 3.0x10−3) before multiple comparison correction. However, it was no more significant after Bonferroni correction. In conclusion, 4 regions out of 7 regions newly reported by the AMD Gene Consortium might be associated with AMD in Japanese.

936W Imputation-Based Genomic Coverage Assessments of Current Human Genotyping Arrays. S.C. Nelson1, K.F. Doheny1, E.W. Pugh2, J.M. Romov3, H. Ilung4, S.R. Browning1, B.S. Weir5, C.C. Laurie1, M. Nishida1,2, K. Kumagai1, K. Yamashiro1, CC. Khor2, CY. Cheng2, TY. Wong2, R. Yamada2, F. Matsuda4, N. Yoshimura1, the Nagahama Study Group. 1) Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Center for Genomic Medicine, Inserm U.852, Kyoto University Graduate School of Medicine, Kyoto, Japan; 3) Fukushima Medical School, Fukushima, Japan; 4) Kobe City General Hospital, Kobe, Japan; 5) Genome Institute of Singapore, Singapore, Singapore.

Genotyping SNPs, combined with imputation of untyped variants, has been widely adopted as an efficient means to interrogate variation across the human genome. The total proportion of genomic variation captured by an array, either by direct observation or imputation, is referred to as ‘genomic coverage’. We have performed imputation-based genomic coverage assessments of eight current genotyping arrays that assay ~0.3 to ~5 million variants. Coverage was estimated using the 1000 Genomes Project phase 1 release, with 1,092 samples representing 14 populations in four continental groups. Samples were divided into ten batches, balancing across populations. In each batch, array variants were used to impute the remaining 1000 Genomes variants, with the rest of the samples serving as the imputation reference. The squared correlation between the observed and imputed allelic dosage (imputation r2) was calculated at each imputed variant over all samples within an ancestry group. The percentage of variants with minor allele frequency (MAF)>0.05 and r2<0.8 is >75% for all arrays and ancestry groups except for African ancestry, and up to ~90% in all ancestries for the highest density arrays. The percentage of variants with 0.01-MAF<0.05 and r2<0.8 is substantially lower: ~40% for low density arrays in all ancestries and 50-80% in high density arrays, depending on ancestry. We calculated genome-wide power to detect variant-trait association in a case-control design, as an average over a grid of MAF by r2 bins weighted by the number of variants per bin and genetic map distance. The most significant SNP was rs8099917 with P = 2.5x10−8 in African ancestry, N = 8,700 (1000 Genomes), 9,700 (Omni2.5M) or 11,700 (HumanCore) in European ancestry and N = 10,100 (1000 Genomes), 11,700 (Omni2.5M) or 19,600 (HumanCore) in Asian ancestry, where N is the total sample size (including equal numbers of cases and controls). These imputation-based genomic coverage and power analyses are intended as a practical guide for researchers weighing the costs and benefits of current array choices in different ancestry groups.

937T Understanding of IL28B gene associated with treatment response for HCV patients. N. Nishida1, Y. Tanaka1, M. Sugiyama2, Y. Mawatari1, M. Iishi1, C. Haga1, K. Tokunaga2, M. Mizokami1 1) Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba, Japan; 2) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Bunkyoku-ku, Tokyo, Japan; 3) Department of Virology & Liver unit, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.

Genome-wide association studies (GWAS) identified single nucleotide polymorphisms (SNPs) near interleukin-28B (IL28B) gene to be associated with virological response (NVR) and virologic response (VR)). Candidate genetic regions with P values below 10−4. Moreover, there were 38 additional NVR and 140 VR identified the strongest associations of rs8099917 with both NVR (P = 9.6x10−5) was calculated at each imputed variant over multiple comparison correction. Since no SNP was genotyped within ±30kb around rs3103783 which is located between IER3 and DDR1, we screened ±50kb region for this SNP. This analysis revealed significant association of this region (lowest P-value = 3.0x10−3) before multiple comparison correction. However, it was no more significant after Bonferroni correction. In conclusion, 4 regions out of 7 regions newly reported by the AMD Gene Consortium might be associated with AMD in Japanese.

To identify novel host genetic factors except for IL28B (rs8099917), we performed a GWAS with increased number of HCV individuals (118 with null virological response (NVR) and 140 with virologic response (VR)). Candidate genetic regions for replication analysis were selected from the result of GWAS by combining the information of gene-gene interactions and genetic pathways. We performed the replication analysis using a total of 895 HCV patients including (321 NVR and 574 VR). A GWAS using Japanese 118 NVR and 140 VR identified the strongest associations of rs8099917 with NVR (P = 9.6 x 10−16, OR=7.62), and also detected 16 novel candidate genetic regions with P values below 10−4. Moreover, there were 56 additional candidate genetic regions to be selected by gene-gene interactions and genetic pathways.

We found no significant association of candidate genes except for IL28B gene, in a total of 895 samples. The combined p-value for rs8099917 reached P = 2.5 x 10−32 (OR=6.00). We also determined the LD structure in a genetic region including IL28B in the Japanese population. Two known associated SNPs (rs1297860 and rs8099917) with HCV drug response had a strong LD with the frame-shift mutation (ss46915590), which located upstream of IFNL3 (IL28B), generated a novel gene, designated IFNL4 (Prokunina-Olsson et al. Nat Genet 2013).

To identify novel host genetic factors except for IL28B (rs8099917), we performed a GWAS with increased number of HCV individuals (118 with null virological response (NVR) and 140 with virologic response (VR)). Candidate genetic regions for replication analysis were selected from the result of GWAS by combining the information of gene-gene interactions and genetic pathways. We performed the replication analysis using a total of 895 HCV patients including (321 NVR and 574 VR). A GWAS using Japanese 118 NVR and 140 VR identified the strongest associations of rs8099917 with NVR (P = 9.6 x 10−16, OR=7.62), and also detected 16 novel candidate genetic regions with P values below 10−4. Moreover, there were 56 additional candidate genetic regions to be selected by gene-gene interactions and genetic pathways.

We found no significant association of candidate genes except for IL28B gene, in a total of 895 samples. The combined p-value for rs8099917 reached P = 2.5 x 10−32 (OR=6.00). We also determined the LD structure in a genetic region including IL28B in the Japanese population. Two known associated SNPs (rs1297860 and rs8099917) with HCV drug response had a strong LD with the frame-shift mutation (ss46915590). Further studies following our report of these robust genetic associations to NVR may make it possible to develop a pre-treatment predictor of which individuals are likely to respond to PEG-IFN-α/RBV treatment. This would remove the need for the initial 12-24 weeks of treatment that is currently used as a basis for a clinical decision about whether treatment should be continued.
Association of body mass-associated polymorphisms in FTO intron 1 with determination of Sasang constitutional types in Koreans. A. Park, S. Cha, H. Yu, J. Kim. Korea Institute of Oriental Medicine, Deajeon, South Korea.

Sasang constitutional (SC) medicine, a part of traditional Korean medicine, classifies humans into 4 constitutional types—Tae-Yang, Tae-Eum (TE), So-Yang, and So-Eum (SE)—based on responses to herbal medicines and individual psycho-physiologic characteristics. Interestingly, the constitutional types are inheritable as revealed by twin and family studies. Recently, several genetic polymorphisms have been associated with SC types from 1,222 subjects via genome-wide association (GWA) analysis. However, the association signal appears to be weak due to the small size of the population used, and the associations are not replicated in the following study. Here, we performed a GWA analysis and replication analysis for SC types determined using an SC analytic tool on the basis of top tertile of the probability values for each type, by scaling up the population size into 5,478 subjects (3,798 subjects in the discovery stage and 1,680 subjects in the replication stage). We found that the minor alleles of polymorphisms in intron 1 of FTO gene were replicably associated with SE type in 2 populations (a peak polymorphism in combined: odds ratio OR = 0.729, p = 1.45E-07). For the other SC types, there were additional associations with the TE type (a peak polymorphism in combined: OR = 1.27, p = 1.45E-05) despite the association signals in the discovery population, but not with the So-Yang type. Since the minor alleles of the polymorphisms in FTO intron 1 have been associated with increased body mass index (BMI), we performed association analyses by controlling BMI. The association signals in SE type remained significant (a peak polymorphism in combined: OR = 0.688, p = 1.28E-04) but not in the TE type. These results corresponded with bodily characteristics of SE and TE types. That is, the subjects with BMI 46 kg/m² associated with BMI-associated polymorphisms in the discovery stage and 1,680 subjects in the replication stage. In conclusion, the BMI-associated polymorphisms in FTO gene were associated well with bodily characteristics of SE and TE types. That is, the subjects with BMI 46 kg/m² associated with BMI-associated polymorphisms in the discovery stage and 1,680 subjects in the replication stage. In conclusion, the BMI-associated polymorphisms in FTO gene were associated well with bodily characteristics of SE and TE types. That is, the subjects with BMI 46 kg/m² associated with BMI-associated polymorphisms in the discovery stage and 1,680 subjects in the replication stage. In conclusion, the BMI-associated polymorphisms in FTO gene were associated well with bodily characteristics of SE and TE types. That is, the subjects with BMI 46 kg/m² associated with BMI-associated polymorphisms in the discovery stage and 1,680 subjects in the replication stage. In conclusion, the BMI-associated polymorphisms in FTO gene were associated well with bodily characteristics of SE and TE types. That is, the subjects with BMI 46 kg/m² associated with BMI-associated polymorphisms in the discovery stage and 1,680 subjects in the replication stage.
Genome-wide association study and admixture mapping reveals new loci associated with total IgE levels in Latinos. M. Pino-Yanes1, C.R. Gignoux2, J.M. Galanter2,3, A.M. Levin2, R. Mathias4, C. Eng1, E.A. Nguy-en1, L.A. Roth1, S. Huntsman5, K. Sandoval1, A. Moreno1, C.A. Winkler6, L.N. Borrell7, B.A. Raby8, S.T. Weiss9, D.L. Nicolae10, G. Ober1, D.A. Mey-ers11, E.R. Bleecker12, F.D. Martinez13,12, S. Sen13, R. Kumar14, C. Busta-mante15, K.C. Barnes15, L.K. Williams15,16, D.G. Torgerson1, E.G. Burchard1,7* On behalf of the GALA II investigators. 1) Department of Medicine, University of California, San Francisco, CA, USA; 2) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA; 3) Department of Public Health Sciences, Henry Ford Health System, Detroit, Michigan; 4) Division of Allergy & Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA; 5) Department of Genetics, Stanford University, Stanford, CA, USA; 6) Basic Research Laboratory, SAIC-Frederick, Inc., Center for Cancer Research, National Cancer Institute, Frederick, MD, USA; 7) Department of Health Sciences, Graduate Program in Public Health, City University of New York, Bronx, NY, USA; 8) Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massa- chusetts, USA; 9) Department of Human Genetics, University of Chicago, Chicago, Illinois, USA; 10) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA; 11) Arizona Respiratory Center, University of Arizona, Tucson, Arizona, USA; 12) Center for Health Policy and Health Services BIOS Institute, University of Arizona, Tucson, Arizona, USA; 13) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA; 14) Children’s Memorial Hospital and the Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA; 15) Department of Medicine, Northwestern University, Chicago, Illinois, USA; 16) Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan.

IgE is the immune system's key mediator of allergic inflammation and is frequently elevated in allergic disorders, such as asthma. Although IgE levels are influenced by the environment, there is an important genetic contribution with estimates of heritability as high as 80%. IgE varies by race-ethnicity in the United States, with higher levels reported in both Latinos and African Americans. The heritability is estimated at 56% in African Americans, 64% in Latinos, and 71% in European Americans. Genome-wide association studies (GWAS) have identified ten genes associated with total serum IgE levels, most have been performed in populations of European descent.

We performed a GWAS and admixture mapping study of total IgE levels in 3,334 Latinos participating in the Genes-environments & Admixture in Latino Americans (GALA II) Study. Our GWAS identified a genome-wide significant association of a common intronic indel in ZNF365 with total IgE (rs2200776616, p=2.3x10^-8), which replicated in an independent sample of Latinos (n=216, p=0.03). We identified four admixture mapping peaks where significant associations of a common intronic indel in MYCBP2 (rs1846190, p=6.1x10^-8) were identified. This mutation is common in Sardinia (MAF=5%) but very rare elsewhere. Genes marked by this mutation are involved in the immune response and can help to elucidate the factors and mechanisms underlying the biological phenomenon. Identifying the genes that influence levels of pro-inflammatory molecules will help to explain the estimated heritability. This study reveals novel loci associated with total IgE levels in Latinos.

Inflammation plays a key role in response to a tissue injury often caused by invading pathogens. It has also been implicated in the pathogenesis of numerous chronic diseases but we are still far from a full understanding of this complex phenomenon. In this study, we measured the levels of the key inflammatory biomarkers Adiponectin, the high-sensitivity C-reactive protein (hsCRP), erythrocyte sedimentation rate (ESR), monocyte chemotactic protein-1 (MCP-1) and interluekin-6 (IL-6) in ~6,000 individuals enrolled in the SardiNIA project. Samples were genotyped with four different Illumina genotyping arrays: OmniExpress, ExomeChip, Cardio-MetaboChip and ImmunoChip and in order to increase the genomic resolution, we imputed low-pass sequencing of 2,120 Sardinians. For each trait, we carried out a sequencing-based genome-wide association scan that allowed us to fine map known signals and detect novel loci. The impact of using sequencing data was clear for all traits, with the exception of IL-6, for which no major differences were seen compared to our previous HapMap GWAS. Interestingly, the same genomic resolution was not reached when using the 1000 Genomes reference panel for imputation. For example, with the 1000 Genomes reference panel we were unable to detect the association of the HBS1L gene for erythrocyte sedimentation rate, with the stop codon variant Q40X. This mutation is common in Sardinia (MAF=5%) but very rare elsewhere. It was present in only one haplotype among 1000 Genomes samples, and thus poorly imputed as rare variant (R2=0.31, MAF=0.008%). Association was missed in the 1000 Genomes imputed data set for three other variants. Consequently, when considering the top variants detected with imputation of the Sardinian haplotypes, the explained heritability was higher for four of five traits than when estimated with the top variants observed after 1000 Genomes imputation. Our study proves the importance of population specific reference panels and their value over the base 1000 Genomes reference set.
Endometriosis is a complex women’s disease with heritability estimated at 52%. A meta-analysis of the published Japanese, Australian and UK endometriosis GWASs confirmed three published loci including, CDKN2A-BAS (rs1537377; \(P=2.4\times10^{-9}\)), intergenic chromosome 7 (rs12700667; \(P=3.6\times10^{-8}\)), WNT4 (rs7521902; \(P=3.2\times10^{-11}\)) and provided evidence for a further four (Nyholt 2012). However, since the publication of the meta-analysis, a fourth GWAS in women of European ancestry from the US was published (Albertsen 2013), involving 2019 cases and 14471 controls. In addition, two replication studies in women of European ancestry were published involving: (1) 1129 cases and 831 controls from Belgium (Sundqvist et al. 2011), (2) 305 cases and 2710 controls from Italy (Pagliardini 2013). Some of the replication evidence for locus-specific associations in these recent papers was non-significant and has generated hypotheses of heterogeneity in the genetic loci underlying endometriosis in different clinical studies even within populations of similar-European-ancestry. Here, our aim is to investigate the heterogeneity and consistency of results across all published studies involving: (1) 1129 cases and 831 controls from Belgium (Sundqvist et al. 2011), (2) 305 cases and 2710 controls from Italy (Pagliardini 2013). Some of the replication evidence for locus-specific associations in these recent papers was non-significant and has generated hypotheses of heterogeneity in the genetic loci underlying endometriosis in different clinical studies even within populations of similar-European-ancestry. Here, our aim is to investigate the heterogeneity and consistency of results across all published eight GWAS and replication studies from Australia, Belgium, Italy, Japan, the UK, and the US through fixed and random effect meta-analyses. The results show robust evidence for nine genetic loci associated with endometriosis, including rs17000667 (\(P<5\times10^{-8}\)) of which association with endometriosis was contested in the two published replication studies-and rs7521902 in WNT4 (\(P=1.4\times10^{-12}\)), with consistent directions of effect across studies and populations and showing no evidence for heterogeneity. Furthermore, six of nine loci are associated with stronger effect sizes among more severe endometriosis cases (raFS 3/4), implying that they are likely to be implicated only in severe disease. Whilst three variants are in intergenic regions and have unknown functionality, the remainder are in/near genes with known functions that are involved in the regulation of reproductive functions. While remaining to be followed up, the results are reassuring in terms of their consistency, they also show that the phenotypic definitions used in the GWAS to date are crude. This emphasizes the importance for future studies to include more detailed phenotypic data, an objective that formed the aim of EPIC Investigators involved in the study. Here, we aimed at harmonisation of clinical phenotype and biological sample collection in endometriosis research (http://endometriosisfoundation.org/epheclt/).
African Population - A Genome-Wide Association Study of Sickle-Cell Uncovering Genetic Modifiers of Sickle Cell Anemia in a Sub-Saharan and inflammation pathways. We have identified several SNPs that may influence susceptibility to S. including the protocadherin-gamma pathway and MAPK signalling pathway. analysis highlighted several biologically interesting candidates involved in showed no evidence of confounding by population stratification. No individual Avg=73.6 y; P=5.70E-04). Subjects with higher BMI and/or type 2 diabetes entailed in the dataset to minimize the likelihood of false positive results. Males were Extensive permutation tests were conducted to obtain appropriate P-values for gene-based and pathway-based results. Additionally, we compared results to those generated under the null by randomly sampling from the entire dataset to minimize the likelihood of false positive results. Males were significantly more likely to be cases than females (P=1.50E-06). The average cas... cases. Moreover, the absence of superimposed relapses and male gender disease is better explained by common genetic variants in BOMS than PrMS 4.8%±1.5% vs 1.7%±0.6%; p=0.05). Our results suggest that the liability of PPMS in males (wGRS: 6.63 vs 6.51, p=0.04; explained variance: further increased (14.7%±6.9%) when considering pure primary progressive days in BOMS (36.5%±10.1%) than in PrMS (20.8%±6.0%), and this difference variance in disease liability explained by 296,391 autosomal SNPs in a characteristic of the onset: progressive onset (PrMS) and bout onset (BOMS) disease. It could be divided in two main clinical courses based on the geographic origin: African and Caucasian population: TNFSF15, IL23R, the major histocompatibility complex (MHC) region, and the RNASET2-GFRFR10P-CRC6 region. Together, the novel and replicated loci accounted for 5.31% of the total genetic variance for CD risk in Koreans. Our study provides new biological insight to CD and supports the complementary value of genetic studies in differ... 94F Identification of 3 Susceptibility Loci for Crohn's Disease in a Genome-Wide Association Study of a Korean Population. K. Song, S. Yang, M. Hong, W. Zhao, J. Baek, I. Liu, Y. Jung, 1) Biochem and Molecular Biol, Univ of Ulsan College of Medicine, Seoul, South Korea; 2) Internal Medicine, Univ Ulsan Col Medicine, Seoul, Korea; 3) Genome Institute of Singapore, Singapore. We have performed a three-stage genome-wide association study in 2,311 Korean patients with Crohn's disease and 2,442 controls. Three new susceptibility loci were identified at genome-wide significance: SNP1 at 4p14 (OR=1.43, Pcombined = 3.60 x 10^-14), SNP2 at 1q25 (OR=1.42, Pcombined = 1.55 x 10^-10), and SNP3 at 1q13 (OR=1.46, Pcombined = 7.15 x 10^-9). Using data from the International IBD Genetics Consortium, disease associations of SNP1 (P=0.00024) and SNP2 (P=5.32 x 10^-5) were replicated in Caucasians. We also replicated four previously reported loci of Caucasian population: TNFSF15, IL23R, the major histocompatibility complex (MHC) region, and the RNASET2-GFRFR10P-CRC6 region. Together, the novel and replicated loci accounted for 5.31% of the total genetic variance for CD risk in Koreans. Our study provides new biological insight to CD and supports the complementary value of genetic studies in different populations.

948W Uncovering Genetic Modifiers of Sickle Cell Anemia in a Sub-Saharan African Population - A Genome-Wide Association Study of Sickle-Cell Severity. T. Singh, S. Nkya, S.E. Cox, J.C. Barrett, J. Makani, 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 2) Muhimbili Wellcome Programme, Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, Tanzania; 3) MRC International Nutrition Group, London School of Hygiene & Tropical Medicine, London, UK; 4) Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK. SACKGROUND: Sickle cell anemia (SCA) is the most common inherited hemoglobinopathy caused by the homozygous inheritance of sickle hemoglobin (HbS) and mostly affects populations of African descent. While the causal variant of the disorder has long been known, SCA results in a heterogeneous set of clinical outcomes affecting all organ systems, and usually includes varying degrees of chronic anemia and episodic acute pain. In our association study, we hope to find common variants that explain the heterogeneity of SCA severity, and in particular, to find further variants that contribute to the heritability of fetal hemoglobin (HBF) levels, which correlates with SCA outcome. This is the largest genome-wide association study analyzing SCA in a population based in Africa, with the widest range of hematological traits and parameters. METHODS: 1167 patients with a phenotype of HbSS or HbS/B, based on quantification of hemoglobin fractions using HPLC (Variant I, Beta-thalassaemia short programme, BioRad, Hercules, CA, USA) from the Muhimbili Sickle Cohort in Dar-es-Salaam, Tanzania were genotyped. Following QC and imputation, associations tests were performed on 1,969,204 hemoglobin (HbF) SNPs typed on the Illumina OmniExpress platform. RESULTS: De novo identified SCA patients in our cohort, substantial population stratification was still observed in our initial association tests. A linear mixed model framework successfully controlled for population stratification, genetic and cryptic relat-edness. Little evidence for genome-wide associations was observed in our cohort, and the dataset, and are in the process of initiating a replication study to verify our results in an independent population with SCA. 950F Genetic burden of common variants in progressive and bout onset multiple sclerosis. M. Sorosina, P. Brambilla, F. Clarelli, N. Banz-zzone, S. Lupoli, C. Guaschino, G. Liberati, A.M. Osicinaru, V. Martinelli, D. Cus, M. Leone, G. Comi, S. D’Alfonso, F. Martinelli Boneschi, International Multiple Sclerosis Genetics Consortium, PRO-GEMUS, PROGRESSO. 1) INSPE, San Raffaele Scientific Institute, Milan, MI, Italy; 2) Interdisciplinary Research Center of Autoimmune Disease IRCAD, University of Eastern Piedmont, Novara, Italy; 3) Department of Health Sciences, University of Eastern Piedmont, Novara, Italy; 4) Department of Health Sciences, University of Milan and Genomics & Bioinformatics Unit, Fondazione Filarete, Milan, Italy; 5) Department of Neurology, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy. Multiple sclerosis (MS) is a neurological autoimmune disease in which genetic and environmental risk factors influence the risk to develop the disease. It could be divided in two main clinical courses based on the characteristic of the onset: progressive onset (PrMS) and bout onset (BOMS) MS. The genetic component of MS has been extensively studied and several loci was found to be associated with the susceptibility to the disease, while the contribution of common genetic variants in the risk of developing the different clinical courses is still unclear. We estimated the proportion of variance in disease liability explained by 296,391 autosomal SNPs in a cohort of Italian PrMS and BOMS patients using the genome-wide complex trait analysis (GCTA) tool and we calculated a weighted log-additive genetic risk score (wGRS) based on the known MS genetic risk variants to evaluate and compare the role of common variants in the disease courses of MS. Our results identified that common SNPs explain more phenotype variance in BOMS (36.5±10.1%) than in PrMS (20.8±6.0%), and this difference further increased (14.7±6.9%) when considering pure primary progressive (PPMS) cases. Similarly, genetic burden measured using wGRS and explained variance by MS-associated SNPs were higher in BOMS than in PrMS in males (wGRS: 6.63 vs 6.51, p=0.04; explained variance: 4.8±1.5% vs 1.7±0.6%; p=0.05). Our results suggest that the liability of disease is better explained by common genetic variants in BOMS than PrMS cases. Moreover, the absence of superimposed relapses and male gender further increase the difference between clinical courses.
951W
Dissecting genomic architecture through advanced SNP-based heritability analysis. D. Speed1, V. Plagnol1, M. Johnson2, D. Balding1. 1) UCL Genetics Institute, University College London, London, United Kingdom; 2) Division of Brain Sciences, Imperial College London, United Kingdom.
SNP-based heritability analysis has had a large impact on the study of complex traits in recent years, but we show that extensions of the method can be used to answer many more questions than have been appreciated to date. We develop several such extensions and apply them to make detailed inferences about the genetic architecture of complex traits. We develop a likelihood ratio test to compare the heritability of different sets of SNPs relative to their sizes (measured in terms of genetic variance). Using this approach, we find that for rheumatoid arthritis and type 1 diabetes, the eQTLs have 50 to 100 times higher intensity of heritability than non-eQTL SNPs, but for other traits eQTLs carry little more heritability than expected by chance. We further find that exonic SNPs explain on average 4.5 times the heritability of non-genic SNPs, and that regions flanking exons have significantly inflated contribution up to 30Kb (after allowing for LD). We compare the heritability of SNPs that are, and are not, associated with a second trait, and use this approach to replicate the known concordance between schizophrenia and bipolar disorder, and to obtain novel evidence of Schizophrenia and type 1 diabetes having a shared genetic architecture. By likelihood-based comparison of different heritability models, we find that although there is a tendency for variants of lower frequency to have greater effect size, this relationship is weaker than commonly assumed. Finally, we find that SNP-based heritability analysis provides a robust method for gene-based analysis, one which is better able to detect causal genes than single-SNP tests of association and that lends itself to subsequent testing of pathological pathways. All analyses are facilitated by our software LDAK.

952T
The human leukocyte antigen (HLA) system is a highly polymorphic region on chromosome 6 that encodes several hundred genes critical for the immune system. HLA has long been known to play an important role in susceptibility and resistance to many infectious diseases and responsiveness to pathogens or vaccines. Many studies have shown significant associations between HLA loci and major infectious diseases, such as HIV/AIDS, hepatitis and malaria etc. However, most studies have tested only candidate loci in small samples, and the associations between HLA alleles and many other infectious diseases are not well studied. We conducted genome wide association studies with over 61,000 individuals who were genotyped on the 23andme platform and asked to report on multiple infectious diseases. We identified single nucleotide polymorphisms in the HLA region that are associated with cold scores, plantar warts, mumps, positive TB test, scarlet fever and shingles at a genome-wide-significant level. We did not detect associations between HLA loci and bladder infection, urinary tract infection, measles, whooping cough, mycoplasma pneumonia or rheumatic fever. To explore the relationship between individual HLA alleles and susceptibility to infectious diseases we used HIBAG, a statistical method combining a large database of individuals with known HLA alleles and SNP variation within the MHC region, to impute HLA alleles at key class I and class II loci over all 23andme customers of European ancestry. We found that HLA-B*18:01 is the HLA allele most significantly (p=3.5e-07) associated, in a protective manner, with cold sores. Individuals with the HLA-DOA1*03:01 allele were less likely to have plantar warts (p=1.7e-08), HLA-DRB1*05 is the HLA allele most significantly associated with mumps (p=1.6e-12). Individuals with the HLA-DRB1*01:03 allele are more susceptible to tuberculosis (p=4.5e-12). Individuals with the HLA-DOB1*03:01 allele are less likely to have scarlet fever (p=9.0e-09). HLA-B*44:02 is an important factor in the resistance to shingles (p=3.4e-11). Although additional studies will be required to separate and validate the association signals in this complex region, combining accurate HLA typing with our knowledge of the function of specific HLA molecules in immune response and with GWAS can be a powerful tool for dissecting infectious disease etiology.

953F
Improving Genome-Wide Association Studies via Markers Decorrelation. O. Weissbrod1, D. Geiger2, N. Zaitlen2, D. Heckerman1. 1) Computer Science Department, Technion - Israel Institute of Technology, Haifa, Israel; 2) Department of Medicine, Lung Biology Center, University of California San Francisco, San Francisco; 3) eScience Group, Microsoft Research, Los Angeles.
Contemporary state of the art genome-wide association studies typically use linear mixed models (LMMs), due to their robustness to false positive results in the presence of population or family structure. It has recently been widely recognized that case-control studies which use covariates are prone to power loss due to ascertainment bias, where the frequency of cases in the study is larger than in the population. Such bias appears in most case-control studies. As we show in this work, LMMs are especially prone to power loss in such cases, because they implicitly treat every single nucleotide polymorphism (SNP) as a covariate. The result is that many risk variants elude detection. We devised a statistical test that bears many similarities to an LMM regarding its ability to properly control the false positive rate, but does not suffer power loss due to ascertainment, thus solving both fundamental problems of association studies. The principal idea behind our method is that true risk variants can be teased apart from false ones via a simple linear (Mahalanobis) transformation for variables decorrelation. After applying the transformation, SNPs that were correlated with the disease due to confounding are no longer correlated. A simple statistical test can then be carried out on the transformed data. We demonstrate that our method outperforms LMMs in ascertainment studies of stratified samples via analyses of synthetic and real data. Simulations suggest that our method can have over 40% power gain over LMMs in realistic scenarios, and over 90% for diseases under extreme ascertainment, corresponding to diseases with 0.1% prevalence. We used our method for analyses of Crohn’s disease, type I diabetes, and Ulcerative colitis from the Wellcome Trust Case Control Consortium, and a highly stratified study of smoking habits from the Genome Analysis Workshop 14 data set. In all cases, our method managed to identify as many or more risk variants than an LMM, with a properly controlled false positive rate. We also verified that our method consistently assigned a statistically significant lower p-value to known risk variants. These results suggest that existing GWAS datasets may be sufficiently powered to reveal dozens of yet undiscovered disease loci.

954W
A group sparse additive model for genome-wide association studies of dynamic complex traits. J. Yin1, M. Marchetti-Bowick2, J. Howrylak3, E. Xing1,2. 1) Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA; 2) Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA; 3) Division of Pulmonary and Critical Care Medicine, Penn State, Milton S. Hershey Medical Center, Hershey, PA.
Despite the widespread popularity of using genome-wide association studies (GWAS) to perform genetic mapping of complex traits, most existing GWAS methodologies are still limited to the use of static phenotypes measured at a single point in time. In this work, we propose a new method for performing association mapping that considers dynamic phenotypes measured at a sequence of time points. Our approach relies on Group Sparse Additive Models (GroupSpAM) for nonparametric regression. This new technique detects a sparse set of genomic loci that are associated with trait dynamics, while simultaneously learning an explicit representation of the dynamic effects of genetic variants at each such locus. We perform a proof-of-concept analysis for detecting single nucleotide polymorphisms (SNPs) associated with the FEV1 score, a sensitive measure of airway obstruction used to assess asthma severity. We evaluate our method using a set of 1500 SNPs from the TGFβ1 gene, which has an established role in asthma pathogenesis and thus serves as a positive control.
955T
Characterizing and redefining clinical subtypes of inflammatory bowel disease (IBD) using genotypes and phenotypes from 47,000 patients. G. Boucher1,2 on behalf of the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). 1) Research Center, Montreal Heart Institute, Montreal, Quebec, Canada; 2) Department of Medicine, Université de Montréal, Montreal, Quebec, Canada.

To date, large-scale genetic studies in IBD have concentrated on the broad clinical diagnoses of Crohn’s disease (CD) and ulcerative colitis (UC). However, it is clear that these conditions are clinically heterogenous and encompass a wide range of subtypes, each with its own pattern of disease behaviour, location and outcome. Whether these subtypes of IBD are distinct diseases with different aetiologies and/or parts of a continuum remains open to debate. Recent progress in characterizing the genetic architecture of IBD, including the identification of 163 IBD-associated loci, affords a unique opportunity to address important clinical questions through the application of genetic risk algorithms. The IIBDGC (a multi-site international consortium) cohort is the largest IBD bioresource generated to date comprising data from over 47,000 patients. Clinical and demographic information on over 26,700 CD and 21,000 UC cases including gender, age at diagnosis, disease location and behaviour, duration of follow-up, surgical history and smoking status were collected according to agreed criteria. We conducted genotype-phenotype analyses across >150k variants (Immunochip), using both established and novel statistical methods for analyses of multinominal phenotypes. We confirmed genome-wide significant associations between the NOD2 and MHC loci and multiple CD phenotypes. We also confirmed an association of the MHC region with extensive UC. Moreover, our analyses implicated multiple additional variants in clinically relevant sub-types. Using the 163 validated IBD loci we showed a significant association between both CD and UC risk scores and CD disease location and behaviour. This association remained highly significant (p-value<10−20) after removing the HLA and NOD2 signals, confirming a role for many IBD risk loci in CD sub-phenotype. Furthermore, we used risk scores to show that colonic CD (a subtype phenotypically similar to UC) is genetically intermediate between ileal CD and UC. In conclusion, our data demonstrate significant genetic influence on IBD clinical heterogeneity. They support the hypothesis that UC, colonic CD and ileal CD could be considered as part of a continuous spectrum. These findings will enhance our understanding of the underlying pathogenesis of the subtypes within IBD, and will likely facilitate a molecular classification of IBD subtypes - a major step on the pathway to personalized medicine.

956F
Genome wide association in 5 isolated populations give new insight on the genetic bases of food preferences. N. Pirastu1, M. Kooyman2, A. Robino3, C. van Duijn4,5, D. Toniolo4,5, P. Gasparini1. 1) Medical Genetics, Institute for Maternal and Child Health - IRCCS ‘Burlo Garofolo’ - Trieste, Univ, Trieste, Trieste, Italy; 2) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam 3000 CA, the Netherlands; 3) Centre for Medical Systems Biology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands; 4) DGCB, San Raffaele Research Institute, Milan; 5) IGM-CN, Pavia.

Food preferences are the first factor driving food choice and thus nutrition. It involves numerous different senses such as taste and olfaction plus numerous other factors such as personal experiences and hedonistic aspects. It is clear that all of these have a genetic base although up to now very limited studies have been conducted. In this work we aimed at identifying the genes that underlie to food preferences using 5 different isolated populations. We have collected during whole population screening ~5000 samples coming 3 Italian isolated population from the INGI network, 1 Western European isolated population from the Netherlands (Erasmus Rucphen Family (ERF) study) plus one coming from isolated communities scattered along 8 countries along the silk road (SR). For each participant we collected numerous information regarding life style, health plus, genome wide genotypes using high density SNP arrays and a food preferences questionnaire which included more than 60 common foods such as liver, orange juice and dark chocolate. Since many of the preferences were specific to each population we decided to restrict our analysis to 22 common foods. We run genome wide association on all 5 cohorts using 1000G imputed SNPs. In order to account for relatedness between samples analyses were conducted using GenABEL and MixABEL. Given the reduced sample size analyses were conducted also on non-additive models (dominant, recessive and overdominant) in order to maximize power to detect associations. Metanalysis was then conducted on the European cohorts and SR was used for replication. Overall we found 19 genome wide significant results for liking of Artichokes (4), Eggplant (2), Whole Milk(2), Dark Chocolate(2), Hot Tea(1), Orange Juice(1), Whipped Cream(2), Red Wine(1), White Wine(2), Mushrooms(1) and Ice Cream (1). Since none of the identified significant SNPs was replicated in the SR cohort we decided to replicate with regional association analysis. We defined around each SNP a 10kb region and estimated region wide significance using PCA-LRT as described in Wang and Abbot 2008. If the SNP fell inside a gene the whole gene was also tested. This type of analysis allowed us to replicate 3 loci one for Artichokes one for Eggplant and one for Whole Milk. This study shows for the first time replicated genome wide loci for liking of common foods and represent a good starting point in understanding the genetic bases of food preferences and consumption in general.
957W

Analysis of Exomic Variation in the Ashkenazi Jewish Population Identifies Novel Associations to Crohn’s Disease in LRRK2. K. Hu\textsuperscript{1}, W. Zhang\textsuperscript{1}, T. Harlütions\textsuperscript{1}, S. Carmi\textsuperscript{1}, B.M. Bowen\textsuperscript{1}, S.R. Brant\textsuperscript{1}, J.D. Rioux\textsuperscript{1}, M. Silverberg\textsuperscript{1}, S. Katz\textsuperscript{2}, A. Chafetz\textsuperscript{2}, H. Zhao\textsuperscript{3}, G. Atzmon\textsuperscript{3}, L. Ozelius\textsuperscript{3}, S. Bressman\textsuperscript{4}, L.N. Clark\textsuperscript{4,13}, I. Pe'er\textsuperscript{3}, T. Lencz\textsuperscript{4}, D.P. McGovern\textsuperscript{4,16}, R.H. Duerr\textsuperscript{5}, J.H. Cho\textsuperscript{1}, I. Peter\textsuperscript{1}.

1) Internal Medicine and Genetics, Yale University, New Haven, CT; 2) Genetics, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Computer Science, Columbia University, New York, NY; 4) Medicine, Johns Hopkins University, Baltimore, MD; 5) Medicine, University of Montreal, Montreal, QC, Canada; 6) Medicine, University of Toronto, Toronto, ON, Canada; 7) Medicine, North Shore-Long Island Jewish Health System, Manhasset, NY; 8) Medicine, Harvard University, Boston, MA; 9) Pathology and Cell Biology, Columbia University, New York, NY; 10) Psychiatry, Hofstra North Shore-LIJ School of Medicine, Manhasset, NY; 11) Medicine, Cedars-Sinai Medical Center, Los Angeles, CA; 12) Medicine, University of Pittsburgh, Pittsburgh, PA.

Background Crohn’s disease (CD) is 4.3-7.7 times more prevalent in the Ashkenazi Jewish (AJ) population compared to non-Jewish European-ancestry populations. The unique demographic history of the AJ population, characterized by an extreme bottleneck, subsequent explosive population growth, and endogamy, has resulted in increased sharing of long haplotypes. We showed previously that common polymorphisms in established CD loci did not account for the AJ population’s increased disease prevalence. In order to identify additional risk loci associated with CD, we performed comprehensive genotyping of 50 AJ individuals with CD and identified 4,277 novel coding polymorphisms of interest, which we combined with markers on the Illumina HumanExome bead chip. Using this custom array, we conducted genotyping in 1,463 CD unrelated cases and 2,770 independent healthy controls, which were genotyped within the extended LRRK2 locus (LRRK2) at the low-frequency coding mutations N2081D and N551K (OR=1.8, P=1.0×10\textsuperscript{-5} and OR=0.66, P=2.0×10\textsuperscript{-8}, respectively). This elaborates our prior CD finding of genome-wide significant association in the LRRK2 region in a previous study of European ancestry individuals, as well as our previously detected linkage peak with a LOD score of 2.08, driven by AJ families. Both mutations were previously linked to Parkinson’s disease (PD), with N2081D conferring a 24% increased risk, while N551K showed a 12% decreased risk in at least one population. Interestingly, LRRK2 G2019S, which is very strongly associated with PD in the AJ population and occurs in the same domain as N2081D, showed no evidence of association with CD. Combined analysis of genotype data from Ashkenazi PD and CD cases, based on AJ population-specific imputation using a reference panel of 128 whole genome-sequenced individuals, suggested additional extensive genetic pleiotropy within the extended LRRK2 locus. Discussion These findings demonstrate the critical role that uncommon alleles play in CD heritability, providing a key link between structure and function. Our results also point towards complex relationships between the genetic risks that can be uncovered by performing a focused study in the Ashkenazi Jewish population.

957T


The immune system’s defense of the body against assault requires a complex interplay of cell-mediated and humoral immunity including a multitude of cell types, antigen processing and presentations systems, cytokines, and inflammatory factors. As a result, teasing out which components underlie each individual’s susceptibility to immunological conditions is a daunting task. Genome-wide analyses studying these conditions can shed light on which pathways play the biggest roles for different types of conditions. We compared the top associations for a diverse set of inflammatory phenotypes including poison oak contact dermatitis, mosquito bite size, and tonsilllectomy (as a proxy for chronic tonsillitis) with each other and with published associations with autoimmune, allergic, and infectious diseases. For inflammatory phenotypes, significant hits near IL21 and GMCSF2 (mosquito bite size) and OSM, LTB4, CXC13, IFGBP3, and IKZF1 (tonsillitis) suggest a more prominent role for cytokines and inflammatory pathways for these conditions than in allergy or infectious disease. There are several overlaps as well. For tonsillitis, the variant near IKZF1 is in high linkage disequilibrium with a variant strongly associated with systemic lupus erythematosus, an autoimmune condition. The region near IL2 and IL21 associated with mosquito bite size also associated with allergy, though the respective variants are not closely linked. In addition, like for many autoimmune, allergic, and infectious conditions, the HLA region comes up as an important hit for both mosquito bite size and tonsilllectomy. In our data, the same variant between HLA-B and MICA (rs130995858) associated with tonsillitis is also associated with anaphylaxis (P=3.4×10\textsuperscript{-8}). For poison oak contact dermatitis, we see a hit within a cluster of CD1 genes on chromosome 1, rather than in the HLA region. CD1 proteins mediate presentation of lipids to T cells (instead of proteins like the major histocompatibility complex). This is the first observed role for CD1 genes, with urushiol, the oily substance produced by the poison oak plant, in triggering the dermatitis. These results highlight the myriad of mechanisms used by the immune system, as well as the subtle similarities and differences in those mechanisms, to react to toxic substances and invaders in the environment.

959F

Joint association analysis of genome-wide human and HIV-1 variation. I. Bartha\textsuperscript{1,2,3}, J. Carlson\textsuperscript{1}, P.J. McLaren\textsuperscript{1}, Z. Brumme\textsuperscript{2}, Ch. Brumme\textsuperscript{2}, R. Harrison\textsuperscript{1}, A. Rauch\textsuperscript{3}, H. Günthard\textsuperscript{1}, M. John\textsuperscript{1}, D. Heckerman\textsuperscript{1}, T.M. Allen\textsuperscript{1,4,5}, C.L. Galíndez\textsuperscript{1}, J. Martínez-Picado\textsuperscript{1,2,3}, V. Müller\textsuperscript{2}, I. Telenít\textsuperscript{1}, J. Felley\textsuperscript{1}, HIV Genome to Genomic Study. 1) School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 2) University Hospital and University of Lausanne, Lausanne, Switzerland; 3) École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 4) University of Lausanne, Lausanne, Switzerland; 5) University of Lausanne, Lausanne, Switzerland; 6) BC Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada; 7) University Hospital Bern and University of Bern Inselspital, Bern, Switzerland; 8) Hospital Universitario Germans Trias i Pujol, Badalona, Spain; 9) Hospital Universitario Germans Trias i Pujol, Badalona, Spain; 10) Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA; 11) Instituto de Salud Carlos III, Madrid, Spain; 12) Hospital Universitario Germans Trias i Pujol, Badalona, Spain.

Background Joint analysis of human genetic and HIV sequence variation has been largely limited to HLA alleles and viral mutations in corresponding epitopes. We performed an unbiased genome-wide search for associations between human SNPs and variation across the HIV proteome. We hypothesized that this non-a priori genome-to-genome analysis would identify and map all host selective pressures on the viral genome.

Methods Human genome-wide genotyping and HIV-1 full-length sequencing data were available for the study. Binary variables were created for each variable amino-acid positions, for every amino acid that was present in at least 3 HIV genomes. Human SNP imputation was performed using 1000 Genomes data as reference. Associations between all SNPs and HIV-1 epitope score were tested by logistic regression under an additive genetic model. For each amino acid that had genome-wide significant association, we searched for independently associated SNPs by iteratively conditioning on the most significant SNP.

Result A total of 1071 patients of European ancestry from 7 cohorts and 5 countries were included in the study. After imputation 6 889 656 SNPs were tested for association with more than 3000 different HIV residues. Highly significant associations (p<1E-11) were observed between SNPs in the Major Histocompatibility Complex (MHC) and multiple amino acids in several HIV-1 proteins. SNPs tagging HLA class I alleles strongly associated with viral variation in CTL epitopes targeted by the corresponding alleles. No significant signals were identified outside the MHC.

Conclusion A non a priori genome-to-genome approach maps associations between host SNPs and HIV genomic variation. We confirm the extensive evolutionary effect that MHC exerts on HIV Gag, Pol and Nef and the lack of common variants with strong effects elsewhere in the host genome on within-host viral evolution.
Shared genetic background for Leprosy an Inflammatory Bowel Disease: chinks in the primary defense against pathogens. E. Fester1,2,10, A. Kresentia Irwanto1,3,4,10, S. Ripke1,6, D. Ellinghaus2, R.K. Weersma2, A. Franke2, J.J. Liu3,4,8,9, IBDGC. 1) Department of Genetics, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands; 2) Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands; 3) Human Genetics, Genome Institute of Singapore, Singapore; 4) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; 6) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 7) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 8) School of Nursing, University of California San Francisco, San Francisco, CA.

Host genetic factors likely influence susceptibility to acquiring HIV-1 upon exposure. However, besides a rare 32-base pair deletion in the CCR5 gene, no replicable genetic associations have been reported. To identify novel genetic factors, we conducted a genome-wide association study (GWAS) of HIV-1 acquisition in 2,004 African Americans and 1,132 European Americans from the Urban Health Study, one of the largest studies of street-recruited injection drug users in North America. Study participants were frequency matched on several important environmental and behavioral risk factors to create an approximate 2:1 ratio of high-risk HIV-negative controls to HIV-positive cases. After genotyping on the Illumina Omni1-Quad BeadChip and applying standard quality control metrics, 1,000 Genomes imputation was conducted using IMPUTE2 with reference to the ALL panel, resulting in the availability of ~8 million single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) > 0.5%. Genotyped and imputed SNPs were tested for association with HIV-1 acquisition in the separate ancestral groups using logistic regression models adjusted for age, gender, survey year, behavioral risk class, and 10 eigenvalues. The ancestry-specific GWAS results were then combined via fixed-effects meta-analysis. Eight genetic loci of interest were identified: 1) loci with statistically significant SNP associations (P<5×10^-8), 6 loci with nominally significant SNP associations (P<1×10^-6), and 1 locus with the top genotyped SNP association (P<1×10^-5). Across the 8 loci of interest, SNPs with P<1×10^-3 were selected for independent replication testing in 1,852 African Americans and 681 European Americans from the Women’s Interagency HIV Study. One chromosome 9p13.2 SNP (P=1.38×10^-4) exceeded the statistical significance threshold for replication (P<3.33×10^-4 based on correction for multiple testing). Several SNPs from 3 other loci on chromosomes 5q31.2, 6p21.32, and 9p24.1 were replicated at nominal significance (minimum P ranging from 1.18×10^-3 to 3.64×10^-3). The GWAS-implicated loci with evidence for replication span genes that have biologically plausible roles in the immune system, some of which have specific functional and regulatory links to protease inhibition and HIV viral DNA integration in host cell DNA. Additional replication testing is underway to further support the statistical evidence for association with HIV-1 acquisition.

Diet and exercise are the first-line treatments for obesity. Since the probability of successful long-term weight-loss (several months) depends on the diet or exercise regimen, clinicians and patients would benefit from development of robust early biomarkers of weight-loss response. Here we report on the first genome-wide association study in placebo-treated subjects pooled from multiple weight-loss trials in order to identify genetic predictors among long-term weight-loss with diet and exercise intervention. DNA samples were available from 355 placebo-treated subjects who also followed a diet which targeted a 25% caloric restriction and an exercise regimen designed according to increasing the patient's current level of physical activity. This cohort had a mean baseline BMI 32.0 (s.d. 4.0) and mean age 49.6 (s.d. 11.9) an overall weight-change of -3.3kg (CI -4.0, -2.7) over the 52 week treatment period. DNA samples were genotyped on the Illumina Human 1M DNA Beadchip and additional SNPs were imputed using imputation approaches with the 1000genomes as reference. Depending on the population there were approximately 6 million markers available for analysis after imputation. After adjustments for multiplicity and undetected stratification, the analysis yielded two significant and three suggestive loci in the White population. The largest signal was at chr19p13.13 with an association signal that still met suggestive thresholds (p-value<5×10⁻7). Another single nucleotide deletion at chr10.109889967 was also associated with a p-value 5.85×10⁻8 with a covariate adjusted mean weight-loss of -13.9kg (CI -17.2, -10.5) for the most extreme genotype group. Genetic prediction on large effective sample size in helping patients to better inform non-responders in order to make trial designs more efficient and improve benefit/risk cost ratios in medical practice. However obesity is a multifactorial process and our data suggest that it is unlikely that any single genetic marker will robustly characterize the response to a diet plus exercise intervention for a large segment of the patient population.

963W Plasma Lipids, Chromosome 11q23.3, and the Risk of Infantile Hypertrophic Pyloric Stenosis. B. Feenstra1, F. Geiler2, L. Carstensen3, P.A. Romiti2, I. Baranowska Körberg3, B. Bedell2, C. Krogh1, R. Fan2, F. Svenningsson1, M. Caggana4, A. Nordenskjold5, J.L. Mills6, J.C. Murray6, M. Melbye7, 1 Dept. of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2 Dept. of Epidemiology, University of Iowa, Iowa City, Iowa; 3 Dept. of Women’s and Children’s Health and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 4 Dept. of Pediatrics, University of Iowa, Iowa City, Iowa; 5 Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA; 6 Wadsworth Center, New York State Department of Health, Albany, New York.

Infantile Hypertrophic Pyloric Stenosis (IHPS) is a common condition in which hypertrophy of the pyloric sphincter muscle layer leads to gastric outlet obstruction. Currently, it is the most common condition requiring surgery among newborns. To search the genome for genetic associations with IHPS, we used reference data from the 1000 Genomes Project for imputation into a genome-wide dataset of 1,001 Danish infantile- and child-onset pyloric stenosis cases and 2,371 disease-free controls. The five most strongly associated loci were tested in replication sets from Denmark, Sweden, and the United States with a total of 1,663 cases and 2,315 controls. We found a novel genome-wide significant locus for IHPS at chromosome 11q23.3 in a region harboring the apolipoprotein (APOAI/C3/A4/A5) gene cluster. The most significant SNP at the locus (odds ratio, 1.59; P = 1.9×10⁻10) is strongly correlated with SNPs previously found to be associated with levels of circulating cholesterol. For these SNPs, the cholesterol lowering allele consistently conferred increased risk. Further investigation using a risk score for IHPS risk, we conducted a functional follow-up study based on umbilical cord blood samples from 46 cases and 189 controls of Danish ancestry sampled from the Danish National Birth Cohort. We found that IHPS risk was inversely related to total cholesterol levels at birth. There was a 64% (95% confidence interval 83%-to-26%; P = 0.005) lower risk per mmol per liter and the 25% of newborns with the lowest levels of total cholesterol were at about five times higher risk compared with the 25% with the highest levels. Infantile Hypertrophic Pyloric Stenosis (IHPS) intervention for a large segment of the patient population.

964T Meta-analysis of genome-wide association studies identifies three new susceptibility loci for intracerebral hemorrhage. G. Falcone1,2,3, C. D. Langefeld1,2, D. W. Reitman4, J. R. Rose5,6,7, International Stroke Genetics Consortium. 1) Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA; 2) Division of Neurocritical Care and Emergency Neurology, Department of Neurology, Massachusetts General Hospital, Boston, MA; 3) Hemorrhagic Stroke Research Group, Massachusetts General Hospital, Boston, MA; 4) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 5) Department of Neurology, University of Cincinnati College of Medicine, Cincinnati, OH; 6) Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest University, Winston-Salem, NC.

BACKGROUND: Intracerebral hemorrhage (ICH) is the stroke type with the worst prognosis and has no established treatment. ICH is classified as lobar and nonlobar based on histopathological differences, with genetic variation contributing to risk of both subtypes. We report the first genome wide association study (GWAS) of this condition. METHODS: The discovery phase included 6 ICH GWASs that enrolled subjects from European ancestry in the US and Europe. Cases were ascertained by neurologists blinded to genotype data and classified as lobar or nonlobar based on brain computed tomography. ICH-free controls were selected from ambulatory clinics or by random digit dialing, depending on the study. Genotyping was completed in Illumina 610 or Affymetrix 6.0. Standard quality controls filters were applied and principal component analysis was implemented to account for population structure. Unobserved genotypes were inferred by imputation to 1000 genomes references panels, with subsequent filtering of single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) <1% or low imputation quality. Association testing was carried out using logistic regression assuming additive genetic effects and including age, gender and principal components in the model. Meta-analyses were performed using the inverse normal method, weighting by sample size. Replication of signals with p-value<5×10⁻7) was pursued in an independent multi-ethnic study utilizing direct genotyping. RESULTS: The discovery phase included 1545 cases (664 lobar and 881 nonlobar) and 1481 controls and identified 4 possible loci: for lobar ICH, chromosome 12q21.1 (rs11179580; MAF=0.22, odds ratio 1.60 (1.28-1.98)) and 8p23.1 (rs1639955; MAF=0.37, OR=1.33, p = 7.1×10⁻7); for nonlobar ICH, chromosome 1q22 (rs2984613; MAF=0.31, OR=1.45, p = 8.1×10⁻8) and 3q13.13 (rs46822240; MAF=0.07, OR=2.04, p = 1.4×10⁻6). Replication included 1256 cases (329 whites, 506 blacks and 421 Hispanics) and confirmed 3 loci: 12q21.1 (p=0.037, meta-analysis p=2.3×10⁻8), 1q22 (p=0.009, meta-analysis p=2.0×10⁻9) and 3q13.13 (p=0.048, meta-analysis p=6.7×10⁻7). Chromosome 8p23.1 did not replicate (p=0.72). CONCLUSION: Three susceptibility loci for ICH were identified, all showing subtype-specificity. These results confirm the presence of biological heterogeneity across ICH subtypes and highlight the importance of ascertaining ICH cases accordingly.

965F Association between Obsessive-Compulsive Disorder (OCD) and Phenotypes in a Genome-Wide Association Study. G. Zai1,2, C. Zai1,2, J. Knight1,2, J.L. Kennedy1,2,3, M.A. Richter1,2,3, 1) Neurogenetics Section, CAMH, University of Toronto, ON, Canada; 2) Department of Psychiatry, University of Toronto, ON, Canada; 3) Institute of Medical Science, University of Toronto, ON, Canada; 4) Department of Psychiatry, Sunnybrook Health Science Centre, Toronto, ON, Canada.

Background: Obsessive-compulsive disorder (OCD) is a chronic and debilitating psychiatric disorder with a strong genetic etiology. A recent genome-wide association study (GWAS) reported interesting candidate gene variants with potential susceptibility to developing OCD. Hypothesis: We hypothesize that genetic variation(s) may be associated with OCD symptom severity using the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) severity score and age at onset (AAO). Method: We investigated 357 individuals with OCD and their nuclear families or matched healthy control (age, gender, and ancestry) for an association of genetic variants and phenotypes including Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) severity score and age at onset (AAO) in a GWAS. GWAS was conducted by the OCF Genetic Collaborative Group. Quality control and analyses were conducted using PLINK and R programs. Results: Several suggestive genome-wide association signals were detected with a p-value of 10E-6 on chromosome 3 and 17q22 for Y-BOCS severity score and chromosome 3 and 9 for AAO. Conclusion: Preliminary findings from our GWAS suggested possible involvement of several regions in OCD symptom severity and age of onset. Further analyses are required to characterize these results.
966W


Uterine leiomyomata, commonly known as uterine fibroids, are benign tumors derived from smooth muscle and fibrous tissue in the uterus, and are the leading cause of hysterectomy in the United States. The lifetime risk for a woman to develop fibroids has been estimated to be as high as 25%. Fibroids tend to grow under the influence of estrogen and shrink when estrogen levels are reduced. The underlying causes of uterine fibroids are not well understood, but twin studies suggest that approximately 55% of the variation in susceptibility to fibroids is genetic. To investigate the genetic factors underlying uterine fibroids, we conducted a genome-wide association study (GWAS) of over 4,000 cases and 12,000 controls of unrelated individuals with European ancestry from the 23andMe cohort. We imputed genotypes against 1000 Genomes reference haplotypes. Participants reported via a web-based survey whether or not they had had uterine fibroids. We report one novel genome-wide association and four suggestive associations. The most significant finding is a variant in the spectrin repeat containing 25 gene previously associated with asthma in individuals of European ancestry. SNPs rs17655581 (P = 8.2 × 10-5) and rs7980829 (P = 2.4 × 10-4) were the most significant. These SNPs are located in intergenic regions uncharacterized for asthma. The third most important genetic signal was found with rs803010 (P = 4.4 × 10-4) located in the promoter of PTGDR, which is a gene previously associated with asthma in individuals of European ancestry. Conclusion: This pooling-based GWAS in French Canadian adult women identified new loci associated with asthma. This study also supported the role of PTGDR in this more homogenous subgroup of asthmatic patients. Further validation is required in independent cohorts of adult asthmatic women of European ancestry.

966F

Familial history of chronic rhinosinusitis predicts more severe disease. L. Mfuna Endami1, A. Filali-Mouhim1, P. Boisvert2, LP. Boulet3, Y. Bossé4, M. Desrosiers4,1. 1) Department of Otolaryngology, CHUM HOTEL DIEU, CRCHUM, Montreal, PQ, Canada; 2) Department of Otolaryngology, Saint-François d’Assise Hospital, Québec, PQ, Canada; 3) Centre de Recherche, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Québec, PQ, Canada; 4) Department of Otolaryngology-Head and Neck Surgery, Montreal General Hospital, McGill University, Montreal, QC, Canada.

Introduction: Genetic association studies have linked chronic rhinosinusitis (CRS) with a number of biologically-plausible genes. However, effects are often small and account for only a portion of CRS cases. Recent evidence suggested that pathogens contributing to CRS development share similar pathways as those revealed by genetic association studies. Cohorts of individuals with CRS may thus contain individuals with genetically-determined and acquired disease, thereby diluting the strength of the genetic signal. We wished to determine whether CRS populations would be enriched for genetic basis to disease by restricting analysis to the sub-group of subjects with a positive family history of CRS. More specifically we tested whether this sub-group was associated with i) a distinct disease evolution and ii) increased strength of previously identified genetic associations. Methods: An existing population of 198 French-Canadian patients with CRS refractory to medical and surgical treatment were screened for incidence of CRS in relatives and impact on population demographics and previously reported genetic association tests. Results: 38% of subjects had a family history of CRS. This group had more severe disease as suggested by an earlier age at onset, an earlier age at first endoscopic sinus surgery and a greater number of medical and surgical interventions. Candidate genes were examined and i) increased strength of previously identified genetic associations. Conclusions: Family history of CRS appears to select for a sub-group of patients with more severe disease suggesting a large genetic component. Also, cohorts of individuals with a lower risk of a genetic component by restricting patient recruitment or analysis to individuals with a family history of CRS may help better identify candidate genes implicated in development of CRS.

967T

Identification of susceptibility genes of adult asthma in French Canadian women. J.C. Berube1, E. Lavoie-Charland1, N. Gaudreau1, L. Sbarr1, C. Henry1, L.P. Boulet2, Y. Bosse1,2. 1) Institut universitaire de cardiologie et de pneumologie de Québec, Quebec, Canada; 2) Department of Molecular Medicine, Laval University, Quebec, Canada.

Introduction: Asthma is a heterogeneous disease in which several susceptibility genes have been identified, but together, those genes explain only a small fraction of the heritability. It was postulated that additional genetic factors may be found by studying subgroups of phenotypically similar asthmatic patients. Objective: The goal of this study is to identify single nucleotide polymorphisms (SNPs) associated with asthma in adult women of European ancestry. Methods: A pooling-based genome-wide association study was performed in 240 allergic asthmatic women and 120 non-allergic asthmatic women. The genotyping was performed in six replicates for both pools using the Illumina HumanOmniExpress BeadChip. The 730,525 interrogated SNPs were ranked based on their likelihood of being associated with asthma using the silhouette score calculated by the GenePool software. The top 20 ranked SNPs plus 18 SNPs ranked among the top 2000 and interrogated SNPs were ranked based on their likelihood of being associated with asthma using the silhouette score calculated by the GenePool software. Results: 21 of the 38 SNPs tested by individual genotyping showed P values lower than 0.05 for association with asthma. SNPs rs17655581 (P = 8.2 × 10-5) and rs7980829 (P = 2.4 × 10-4) were the most significant. These SNPs are located in intergenic regions uncharacterized for asthma. The third most important genetic signal was found with rs803010 (P = 4.4 × 10-4) located in the promoter of PTGDR, which is a gene previously associated with asthma in individuals of European ancestry. Conclusion: This pooling-based GWAS in a population of French Canadian adult women identified new loci associated with asthma. This study also supported the role of PTGDR in this more homogenous subgroup of asthmatic patients. Further validation is required in independent cohorts of adult asthmatic women of European ancestry.
969W

Refinement of whole-body bone mineral density measures of children assists the identification of genetic variants associated with skeletal site specificity and bone mass attainment. J.P. Komp et al., C. Medina-Gomez et al., K. Estrada et al., B. St-Pourcain et al., D.H. M. Happe et al., N.J. Timpson et al., L. Oei et al., S.M. Ring et al., C.J. Kruithof et al., L.E. Wolter et al., F.M.K. Wilkins et al., M.C. Zilkens et al., A. Hofman et al., A.G. Uitterlinden et al., G. Davey-Smith et al., V.W.V. Jaddoe et al., J.H. Tobias et al., F. Rivadeneira et al., D.M. Evans et al. 1) MRC Centre for Causal and Translational Epidemiology, University of Bristol, Bristol, United Kingdom; 2) School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 3) Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands; 4) The Generation R Study Group, Erasmus University Medical Center, Rotterdam, The Netherlands; 5) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 6) Netherlands Genomics Initiative (NGL)-sponsored Netherlands Consortium for Healthy Aging (NCHA), The Netherlands; 7) Department of Paediatrics, Erasmus University Medical Center, Rotterdam, The Netherlands; 8) Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK; 9) School of Clinical Sciences, University of Bristol, Bristol, UK. The heritability of bone mineral density (BMD) varies across skeletal sites, reflecting different relative contributions of environmental and genetic influences. To quantify to the degree which common genetic variants and environmental factors influence BMD at different sites, we estimated the genetic (r_g) and environmental (r_e) correlations of BMD measured at the upper limbs (UL), lower limbs (LL) and skull (S), using whole-body DXA scans of ~4890 participants recruited by Avon Longitudinal Study of Parents and their Children (ALSPAC). Point estimates of r_g indicated that appendicular sites shared a greater proportion of shared genetic architecture (UL:UL-BMD r_g=0.78) when compared to the skull (UL/S-BMD r_g=0.58 and LL/S-BMD r_g=0.43). Likewise, environmental factors influencing BMD at the appendicular sites were broadly similar to each other (r_e=0.55), however they were appreciably different from the factors influencing S-BMD (r_e=0.20 - 0.24). To explore the basis for the observed difference in r_g and r_e, genome-wide association meta-analyses were performed (n=9395), combining data from ALSPAC and the Generation R study (GEN-R) to identify genetic signals associated with BMD. We discovered a new paradigm that multiple associated variants differed in the strength of their association and magnitude of effect with each sub-region. In particular, effect sizes of variants situated closest to EYA4 (6q23.2), COLEC10 (8q24.12), LIN7C (11p14.1) and TMF8SF1A (18q21.33) appeared to be larger for S-BMD when compared to UL-BMD. Furthermore variants at the WNT16 and CENPW showed considerable site-specificity as indicated by very strong levels of association with S-BMD and/or UL-BMD but not with LL-BMD. In addition, we report a novel association between a variant in RIN3 and LL-BMD (rs754388; β=0.13, SE=0.02, P=1.4x10^-14) and highlight its prior association with Paget’s disease. We suggest that different skeletal sites as measured by whole body-DXA are to a certain extent under distinct environmental and genetic influences. Allowing for these differences may help to uncover new genetic influences on BMD, particularly those involved in bone accrual, for which S-BMD appears to be particularly well suited.

970T

GWAS SNPs impact gene expression through inheritance of multiple enhancer variants. O. Corradin et al., A. Salakhova et al., B. Akhtar-Zaidi et al., L. Myeroff et al., J. Wiltsie et al., S. Markowitz et al., P.C. Scacheri et al. 1) Department of Genetics and Genomic Sciences, Case Western Reserve University, Cleveland, OH 44122; 2) Department of Pathology, Case Western Reserve University, Cleveland, OH 44122; 3) Department of Medicine, Case Western Reserve University, Cleveland, OH 44122; 4) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106. Since 2005 over 1350 genome-wide association studies (GWAS) have been published, identifying thousands of SNPs associated with more than 600 traits. More than 90% of these SNPs are located outside protein coding regions, and are hypothesized to influence gene expression by disrupting the function of non-coding regulatory elements such as enhancers. Consistent with this hypothesis, the ENCODE consortium and other groups have discovered thousands of GWAS variants to enhancer elements identified through epigenomic profiling studies. However, the genes whose expression these enhancer variants influence and their effect on target transcript levels often remains unknown. Moreover, because GWAS-loci often contain multiple SNPs in linkage disequilibrium (LD) with the ‘lead’ SNP, any one of which may be ‘causative’, pinpointing the culprit is often a major challenge. We provide evidence that for six common autoimmune disorders (rheumatoid arthritis, Crohn’s disease, celiac disease, multiple sclerosis, lupus, and ulcerative colitis), the GWAS association arises from a group of polymorphisms in LD that map to multiple enhancer elements active in the same cell type. This finding suggests that for common traits, there is not necessarily a single disease-causing variant that underlies the association signal, but rather several variants that impact multiple enhancers and cooperatively influence gene expression. We call this the ‘multiple enhancer variant’ hypothesis. Using a novel method to delineate enhancer-gene interactions, we show that multiple enhancer variants within a given locus typically target the same gene. Using available data from HapMap and B lymphoblasts as a model system, we provide evidence at numerous loci that multiple enhancer variants cooperatively contribute to altered expression of their gene targets. The effects on target transcript levels can be either gain or loss of function, and overall tend to be modest. Overall, the multiple enhancer variant hypothesis offers a new paradigm by which non-coding variants confer susceptibility to common traits.

971F

Novel Genetic Association of Primary Severe Localized Provoked Vulvodynia with TRPV1 and NGF: Possible Common Predisposition with Other Pain Syndromes. T. Falik-Zaccal et al., L. Kafoni et al., A. Azran et al., Y. Farajun et al., E. Tubin et al., O. Hemo et al., L. Abramov et al., A. Yeshaya et al., J. Bornstein et al. 1) Institute of Human Genetics, Western Galilee Medical Center, Nahariya, Israel; 2) The Galilee Faculty of Medicine, Bar Ilan, Safed, Israel; 3) Department of Obstetrics & Gynecology, Western Galilee Medical Center, Nahariya, Israel; 4) Lis Women’s Hospital, Tel Aviv, Israel; 5) Snaider Women’s Hospital, Bellinson Medical Center, Petah-Tikva ,Israel; 6) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. Localized Provoked Vulvodynia (LPV) is a multifactorial syndrome affecting 10-16% of women worldwide, causing pain of variable nature and intensity at every attempt to have sexual intercourse. Familial occurrences of LPV suggest genetic susceptibility in the etiology of this condition. We have studied possible associations between LPV and several SNPs in genes previously hypothesized to be involved in the pathophysiology of LPV by coding for proteins involved in mast cells degradation and hyper-innervation in the vestibule of affected women. We have analyzed mixed single nucleotide polymorphisms (SNPs) in 70 women presenting severe primary LPV and 132 healthy, ethnically matched controls. Each participating women have answered a detailed questionnaire, addressing possible familial occurrence of LPV and the existence of pain co-morbidity conditions. SNP analyses were performed for the genes Transient Receptor Potential Vanilloid type-1 (TRPV1), the Nerve Growth Factor (NGF) and the Heparanase (HPSE), using multiplex PCR and SNapShot assay and by restriction fragment length polymorphism. We found a significant difference in familial occurrences of LPV, Temporomandibular joint (TMJ) symptoms, recurrent vaginitis and irritable bowel syndrome (IBS) between affected women and the healthy controls. Novel statistically significant differences in the prevalence of non-synonymous polymorphism c.945C>G (rs222747) of TRPV1 and of SNP localized in the promoter region of NGF (rs1102930) were identified between the affected women especially from Ashkenazi Jewish ancestry and the control group. Data was also analyzed to compare genetic variation with a variety of pain condition among the 202 women. Interestingly, rs222747 minor allele of TRPV1 was found in association with women presenting TMJ and women with recurrent vaginitis, suggesting possible common genetic predisposition to pain co-morbidities. Our results imply possible genetic susceptibility to LPV associated with specific alleles in the genes TRPV1 and NGF, and propose a genetic predisposition for LPV and other pain syndromes through a yet unknown biological pathway involving TRPV1.
972W
Predicting regions associated with complex traits using Multi-kernel Support Vector Machines. D. Kostka1, J.A. Capra2. 1) Developmental Biology and Computational & Systems, University of Pittsburgh, Pittsburgh, PA; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Genome-wide association studies (GWAS) have identified thousands of genomic loci associated with hundreds of complex traits in different populations. However, these associations likely represent only a fraction of the loci contributing to the genetic basis of most complex traits. To aid the search for additional loci and improve our understanding of known associations, we developed an algorithm to distinguish genomic regions containing SNPs associated with a trait from regions associated with other traits and from the genomic background. For 59 traits with more than 40 unique associated SNPs in the NHGRI GWAS Catalog, we computed a tight linkage disequilibrium (LD) block (using appropriate 1000 Genomes populations) for each trait-associated SNP and then characterized these blocks’ evolutionary histories, sequence motif patterns, and overlapping functional genomics data (from ENCODE and Roadmap Epigenomics). We then trained a multiple kernel support vector machine to distinguish regions associated with each trait from those associated with other traits and regions with no known associations. The algorithm obtains strong performance at distinguishing regions for many traits from regions with no associations (ROC AUCs above 0.8), and it performs significantly better than random for all but five traits. Our approach is also able to distinguish among regions associated with different traits; however, there is considerable variation in performance: a maximum ROC AUC of 0.78 for celiac disease and a minimum of 0.50 for adverse metabolic events. Using the trained classifiers for each trait, we predicted thousands of loci across the human genome that resemble known trait-associated regions (at a 5% FDR). The predicted regions for nearly all traits are enriched for nearby annotations and gene expression relevant to the trait. In addition, traits known to share common pathways overlap in their predicted regions more than less related traits. These genome-wide predictions can extend our knowledge of the genetic basis for complex traits. To illustrate this potential, we show how they can aid in prioritization of SNPs with low p-values that do not reach the genome-wide significance level. We also demonstrate the use of our predictions in developing hypotheses about the mechanisms behind these associations by identifying several hundred known enhancers and eQTLs that fall in novel predicted trait-associated regions.

973T

Background: Several lines of evidence point to overlapping molecular mechanisms between late-onset Alzheimer disease (AD) and age-related macular degeneration (AMD), including the molecular composition of the extra-cellular deposits which are the hallmarks of each, shared cardiovascular risk factors, and significant risk associated with APOE in both disorders.

Methods: We evaluated summary statistics from large-scale genome-wide meta-analyses of AD (Naj et al. 2011) and AMD (Fritsche et al. 2013), treating the AD GWAS as a discovery dataset and the AMD dataset as a replication dataset to focus our search for genetic risk factors of AMD at the SNP, gene (significance based on peak SNP adjusted using the Li and Ji method), and pathway level. Results: First, we examined all SNPs associated with AD at a 10% false discovery rate (FDR) cutoff–excluding regions previously associated with AMD. None of the 226 SNPs tested for association with AMD were significant at the .05/226=.0002 level. Next, we examined all putative AD genes significant at a 10% FDR level (excluding the APOE region) for association with AMD. Two of 24 genes examined were significant (p=.05/24=.0021). The gene PILRA on chromosome 7q22 (FDRadj=.057) was associated with AMD (p=7.0E-6; peak SNP rs7792525, MAF=19%, OR=1.4, p=2.3E-6). Second, the known AD risk gene ABCA7 (FDRadj=.0006) was also associated with AMD (p=0.0012, peak SNP rs3752228, MAF=5% OR=1.22, p=1.2E-4). The AMD-associated ABCA7 SNP is not in LD with the previously identified AD risk variants rs3784650 and rs3752246 (r2<.01). Finally, we examined association of AMD with genes in the clathrin-mediated endocytosis signaling pathway (191 genes) and the atherosclerosis signaling pathway (132 genes), which emerged from an INGENUITY canonical pathway analysis of AD risk genes. Four of these genes were significantly associated with AMD including several previously established AMD loci. Novel associations for AMD were observed with HGS (p=7.5E-5; peak SNP rs8070488, MAF=23%, OR=.91, p=1.8E-5), which encodes a protein regulating endosomal sorting, and gene TNF (p=.12E-5; peak SNP rs2071590, MAF=.34%, OR=.89, p=9.8E-7), a proinflammatory cytokine. LD in the regions around TNF and PILRA makes it impossible to localize the association to those genes specifically. Conclusions: We identified several novel modest-effect AMD risk variants in genes with a direct or peripheral role in AD pathogenesis.

1) ECLAMC (Latin American Collaborative Study of Congenital Malformations) INAGEMP (National Institute of Populational Medical Genetics) at Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil; 2) ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP-CNpq (National Institute of Population Medical Genetics), Rio de Janeiro, Brazil; 3) ECLAMC at Center for Medical Education and Clinical Research (CEMIC-CONICET), Buenos Aires, Argentina; 4) Bioinformatic Department, Research Center Principe Felipe, Valencia, Spain; 5) ECLAMC and INAGEMP at Hospital Zonal El Bolsón, El Bolsón, Argentina; 6) División de Genética, Instituto Nacional de Cáncer, Rio de Janeiro, Brazil; 7) Departments of Oral Biology and Pediatric Dentistry and Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh.; 8) ECLAMC and INAGEMP at Congenital Malformations Epidemiology Laboratory, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil.

Cleft lip with or without palate (CL±P) are one of the most common birth defects, worldwide. Its birth prevalence ranges from 1/500 to 1/2000, depending on geographical origin, and ethnic background. The etiology of CL±P is complex and very heterogeneous with multiple genetic and environmental factors playing a role. The identification of genes that participate in the complex etiology of CL±P is an important step considering the morbidity that accompanies this relatively common birth defect. In this scenario, genome wide association studies are important tools in the search for candidates genes associated with complex diseases. Among the candidate genes described to be associated with CL±P some of them are on the X chromosome. Recently two genes (OFD and DMD) were associated with isolated CL±P in Asians and European populations. In this study we selected a CL±P high prevalence population from Patagonia identified by the ECLAMC and 99 individuals were until now ascertained and studied. SNPs from chromosome X were genotyping using the Affymetrix Platfora (Genome-Wide 6.0) and tested (TDT) for association with isolated CL±P in 24 trios, 2 duos, and 23 sibs using PLINK. We found statistical evidence of association with CL±P in 57 SNPs localized in six genes KAL 1 (06), DMD (04), SYN1 (02), CFP (02), TBL1X (01), TLR8 (01), uncharacterized region LOC100873065 (01) and intergenic regions (40) (P<0.008151). To assess the role of the 57 SNPs on CL±P phenotype, we performed haplotype analysis with Haplovap Software. Fourteen haplotypes under P<0.0068 were identified whereas the most significant involves the following SNPS rs16988033-rs7055843-rs16997889, all of them in intergenic regions on locus Xp21.3 (P= 0.0009). The results for DMD and KAL1 genes are consistent with recent data about the involvement of these genes in the etiology of oral clefts while the other genes were described for the first time. In order to confirm the current association and also to improve our statistical power, it’s advisable increase the number of trios affected by CL±P in further studies.
Gene-environment interactions between obesity gene variants and body size in multiple sclerosis. M.A. Gianfrancesco1,2, A. Acuna2, L. Sheni2, F.B.S. Briggs1, H. Quach1, A. Bernstein1, A.K. Hedstrom4, I. Kockum1, L. Alfredsson5,6, T. Otsson5, C. Scaher6, L.F. Barcellos7,2,3,1. 1) School of Public Health, Dept of Epidemiology, UC Berkeley, Berkeley, CA; 2) Kaiser Permanente Division of Research, Oakland, CA; 3) Palm Drive Hospital, Sebastopol, CA; 4) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 5) Neuroimmunology Unit, Department of Clinical Neuro- 6) Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 7) Division of Epidemiology, Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA.

Multiple sclerosis (MS) is a demyelinating autoimmune disease affecting over 2.5 million people worldwide. Although studies have confirmed a strong genetic component, environmental factors, as well as obesity, have been identified as potential contributors to disease risk. In this study, we aimed to investigate genetic and environmental interactions in 1,223,314 SNPs using the Illumina HumanOmni2.5-8 platform. Multipoint linkage analyses in families with early onset of disease were performed for several established quantitative traits, including axial length (AXL), central corneal thickness (CCT), and body mass index (BMI). A total of 92,157 variants in African Americans (~34%) and 84,259 (~31%) in Hispanics were polymorphic and passed quality control metrics. Genotyping was performed using the Illumina HumanExome BeadChip. The highest LOD score was 3.91 with the APOE SNP rs7412 (MAF=0.13) and plasma levels of apolipoprotein B (ApoB). In Hispanics, one of the strongest signals of linkage was observed for LDL cholesterol (LDL_c), accounting for up to 4.5% of the variance in HDL. This study highlights the importance of considering genetic and environmental factors in the development of MS.

Genetic linkage analysis using large consanguineous pedigrees from South India suggests new loci for ocular quantitative traits. B. Fan1,2, P. Ferdina Marie Sharmila1, N. Soumitra2, S. Srinivasa2, J. Madhavan2, D.S. Friedman3, L. Vijaya1, J.L. Haines4, R. George5, J.L. Wiggs1, 1) Department of Ophthalmology, Harvard Med Sch, Massachusetts Eye & Ear Infirmary, Boston, MA; 2) National Research Foundation, Sankara Nethralaya, Chennai, India; 3) Johns Hopkins Medical School, Wilmer Eye Institute, Baltimore, MD; 4) Medical Research Foundation, Sankara Nethralaya, Chennai, India; 5) Center for Human Genetic Research, Vanderbilt University School of Medicine, Nashville, TN.

Many ocular quantitative traits are highly heritable. Mapping genes contributing to these traits can provide insights into disease mechanisms and potential therapeutic targets. In this study, we used genome-wide linkage analyses in families with early onset of disease to identify new loci for ocular quantitative traits. We genotyped 1,223,314 SNPs in 1,204 consanguineous families from South India. Linkage analysis revealed several new loci for ocular traits, including axial length (AXL), central corneal thickness (CCT), and body mass index (BMI). These findings have important implications for the development of new therapeutic strategies for diseases affecting the eye.
979T

Genome-wide screen for self-reported physical disability loci in the oldest-old Amish. J.E. Hicks1, J.R. Gilbert1, L. Caywood1, L. Reinhart-Mercer1, D. Fuzzell2, R. Lau2, M.A. Pericak-Vance1, J.L. Haines1, W.K. Scott1. 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Center for Human Genetics Research, Vanderbilt University, Nashville TN.

As the population ages, age-related disability will increasingly become an important public health concern. However, many objective measures of aging differ from self-reported measures of impairment. To identify patterns of perceived disability in the oldest old, a latent class analysis was performed on four scales assessing perceived disability in 263 cognitively intact participants aged 80 or above. Participants were enrolled through a population-based door-to-door study of community-dwelling Amish individuals aged 65 and older from Indiana and Ohio. Individuals were placed into a 13-generation pedigree using the Anabaptist Genealogy Database. These individuals were assessed with the Activities of Daily Living scale (basic self-care tasks such as dressing and bathing), Instrumental Activities of Daily Living scale (higher-level tasks such as cooking and shopping), the Nage scale for musculoskeletal function, and the Rosow-Breslau scale for lower-extremity function. Data analysis was performed with the pQCA package in R with age and sex included as covariates. The optimal model identified two classes. The first (n=209) was largely unimpaired on all disability questions, while the second class (n=54) reported some degree of physical disability. Random forest analysis of disability found that the strongest predictor of class was ability to perform heavy household. A genome-wide screen was performed with Affymetrix Genome-Wide Human SNP Array 6.0. PLINK was used to select a set of 4,966 SNPs in linkage equilibrium (pairwise r2 < 0.16) for linkage analysis. PedCut created 11 subgroups suitable for linkage analysis, each containing at least 2 of the 54 physically impaired individuals. Multipoint nonparametric linkage analysis (LOD*) was performed in MERLIN and identified no regions of significant linkage to disability but did identify two regions providing suggestive evidence of linkage on chromosomes 1 (LOD*=2.43, 18,922,283 Mb) and 3 (LOD*=2.28, 30,502,223 Mb). While tests of association in the 2-LOD-down support intervals (SI) surrounding these peaks did not identify markers significantly associated with disability when corrected for multiple testing (p<10^-8), they were suggestive of linkage. Analysis of VWF identified a novel locus at 2p12-2q13 (19.2% variance explained) that was significant after correction for multiple testing (p<10^-12) at 12p13, overlapping the VWF gene itself. This SNP encodes a His484Arg variant in the VWF propeptide domain. However, interference of His484Arg with antigen detection was suggested after a repeat GWAS in this cohort using VWFpp as the phenotype. While VWF levels are correlated with VWF clearance rates, we provide mechanistic insight into a previous GWAS of VWFpp. VWF synthesis/secretion rates, we provide mechanistic insight into a previous GWAS of VWF.

980F

Linkage and association analysis of von Willebrand factor propeptide levels provides mechanistic insight into the genetic control of plasma von Willebrand factor (VWF) levels. K.C. Desch1, A.B. Ozen1, D. Siemieniacka1, J.Z. Li1, D. Ginsburg2,3,4. 1) Pediatrics and Communicable Disease, University of Michigan, Ann Arbor, MI; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Howard Hughes Medical Institute, Ann Arbor, MI; 4) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Von Willebrand Factor (VWF) is a plasma glycoprotein playing a central role in hemostasis. Common alleles at the ABO locus account for 24.5% of VWF variance and operate through altered VWF clearance rates. The VWF propeptide (VWFpp) is secreted into circulation with mature VWF where unlike VWF, it is rapidly cleared from circulation. Therefore plasma VWFpp levels are largely determined by the rate of synthesis/secretion of VWF. To gain mechanistic insight into genetic determinants of VWF levels, we performed the linkage and genome-wide association (GWAS) of VWFpp in two healthy cohorts (combined n=3462) and compared the results to our previous analyses of VWF in these cohorts. (Desch et al., 2013) VWFpp levels were determined with monoclonal antibodies in an AlphaLISA assay. 800K SNPs, that were available after extensive QC in our previous study, were used in association studies (PLINK, EMMAX, METAL) and linkage analysis (MERLIN). Heritability (h2) of VWFpp was 74% compared to 65% for VWF. Our previous meta-analysis GWAS of VWF revealed signals at ABO (P=7.9E-139) and VWF (P=5.5E-16). Here, preliminary GWAS for VWFpp in one cohort revealed a single significant SNP rs1800378 (p-value = 8.9E-13) at 12p13, overlapping the VWF gene itself. This SNP encodes a His484Arg variant in the VWF propeptide domain. However, interference of His484Arg with antigen detection was suggested after a repeat GWAS in this cohort using VWFpp as the phenotype. While VWF levels are correlated with VWF clearance rates, we provide mechanistic insight into a previous GWAS of VWF pp. VWF synthesis/secretion rates, we provide mechanistic insight into a previous GWAS of VWF pp.

981W

Genetic variants on 17q23.2 and 10q11.21 are associated with variation in telomere length: The Long Life Family Study. J.H. Lee1,2,3, R. Cheng1, L.S. Hong1,2, M. Feilosa1, C. Kammerer3, M.S. Kang1, N. Schupf1,2, R. Lin1, J.L. Sanders2, H. Bae1, T. Drukey1, T. Perls3, K. Christensen4, M. Provenza1, R. Mayou2,3,1. 1) Sergievsky Ctr/taub Inst, New York, NY; 2) Depts of Neurology and Psychiatry, Columbia University, New York, NY; 3) Dept of Epidemiology, Columbia University, New York, NY; 4) Div of Statistical Genomics, Depr of Genetics, Washington University, St. Louis, MO; 5) Dep of Pediatrics & Genetics, Washington University, St. Louis, MO; 6) Dep of Epidemiology & Center for Aging and Population Health, University of Pittsburgh, Pittsburgh, PA; 7) Div of Geriatrics, Dep of Medicine, Boston University Medical Center, Boston, MA; 8) The Danish Aging Research Center, Epidemiology, University of Southern Denmark, Odense, Denmark.

Telomere length is hypothesized to measure cellular aging in humans, and short telomere length is associated with increased risks of late onset diseases, including cardiovascular disease, dementia, cancer, and other age-related disorders. Many studies have shown that telomere length is a heritable trait, and several candidate genes have been identified, including TERT, TERC, OBFC1, and CTC1. Most studies have focused on genetic causes of chronic diseases such as heart disease and diabetes in relation to telomere length. This investigation involved a genome wide family-based association analyses that identified genetic variants that contribute to variation in telomere length among families with exceptional longevity. From the genome wide association analysis in 4,239 LLFS participants, we identified a novel intergenic SNP rs7680468 near PAPSS1 on 4q24 (p=2.3E-13). In our linkage analysis, we identified two additional loci with HLOD scores>3, including one at 17q23.2 (HLOD=4.77) and another at 10q11.21 (HLOD=4.36). Subsequent gene-wise association analysis identified multiple candidate genes from these two regions, including DCAF7, POLG2, CEP95, and SMURF2 at 17q23.2; and RAGSF1A, HNRPF, ANF,487, CTF2T, and PRKAT1 at 10q11.21. Among these genes, we identified one contiguous haplotype in CEP95 and SMURF2 that was significantly associated with telomere length. In addition, previously reported genes - TERC, ARP1, MYNN, OBFC1, and ZNF729 - showed weak to modest association with telomere length variation. Here we report novel genetic variants and confirm previously reported genes that are associated with variation in telomere length. These may reveal additional insight into cellular aging.
Combined genome-wide linkage and association studies of centenarians identifies several new candidate genes for longevity. P. Sebastiani\textsuperscript{1}, H. Bae\textsuperscript{1}, T.T. Perls\textsuperscript{2}. 1) Dept Biostatistics, Boston Univ Sch Public Health, Boston, MA; 2) Dept Medicine, Boston University School of Medicine, and Boston Medical Center, Boston MA.

The genetic basis of extreme longevity increases as more extreme lifespans are examined. However, genome-wide association studies (GWAS) of longevity have failed to discover highly significant variants. Several studies have confirmed a role of APOE in lifespan and rs2075650 in TOMM40/APOE is the only single nucleotide polymorphism (SNP) that reached genome-wide significance in several GWAS of nonagenarians and centenarians. In a GWAS of 801 centenarians (median age at death 104 years) versus population controls, we showed that a set of 281 SNPs that did not reach genome-wide significance could accurately distinguish centenarians from controls, with increasing sensitivity in older and older people. SNPs that failed to reach genome-wide significance but showed consistent effects in different studies have also been published. These findings suggest that the genetic makeup of exceptional longevity is determined by a combination of common and rare variants with modest genetic effects, and that the relatively small sample size of centenarian studies is a serious limitation to their discovery power. We hypothesized that linkage analysis could be used to prioritize variants that reached statistical significance in association studies but failed to reach genome-wide significance. We conducted a non-parametric linkage analysis of affected sib-pairs enrolled in 172 families of the New England Centenarian Study (NECS) enrolled between 1995 and 2010. Affected pairs were defined as two siblings with age at death exceeding a fixed percentile survival from sex and birth year matched life table from the US Social Security Administration. Percentile survivals ≥ between 0.25 and 0.001 were used to derive nested sets of sib-pairs that were analyzed in Merlin. Regions in chromosomes 2, 3, 4, 7, 10 and 19 had linkage peaks with LOD score >2.4 for at least one threshold σ, or LOD score >2 for more than one threshold σ and 1.5 LOD support intervals were annotated by significant results from two GWAS of longevity in NECS centenarians. The region under the linkage peak in chromosome 10 includes ADARB2 that we found previously associated with extreme longevity. The region in chromosome 3 includes SNPs in MYTL1 and its regulatory region that is enriched of SNPs associated with insulin resistance, insulin variability and diabetes complications. Three regions in chromosome 19 point to additional genes that are associated with extreme longevity independent of APOE.

Analyzing patterns of IBD sharing in Oceanic Palau to identify genomic regions harboring risk for schizophrenia. C.A. Bodea\textsuperscript{1}, F. Middleton\textsuperscript{2}, L. Klei\textsuperscript{3}, S.V. Faraone\textsuperscript{3}, Š. Vinogradov\textsuperscript{3}, J. Tiobrech\textsuperscript{3}, V. Yarno\textsuperscript{3}, S. Kuarter\textsuperscript{3}, K. Roeder\textsuperscript{4}, N. Melhem\textsuperscript{2}, B. Devlin\textsuperscript{3}, M. Myles-Worsley\textsuperscript{2}, W. Byerley\textsuperscript{4}. 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 2) Department of Psychiatry, SUNY Upstate Medical University, Syracuse NY; 3) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 4) Department of Psychiatry, University of California San Francisco, Pittsburgh, PA; 5) Palauan Ministry of Health, Republic of Palau, Pittsburgh, PA.

The genetics of schizophrenia is complex and, despite recent successes, it remains challenging to identify individual risk loci. We have been studying the genetics of schizophrenia in Oceanic Palau, an island nation some 500 miles east of the Philippines. One of the features we have been analyzing is the co-segregation pattern of genetic variation and schizophrenia status to identify regions of the genome shared identical-by-descent (IBD) among related, affected individuals. These IBD segments are shared as haplotypes and we search for inherited haplotypes that increase the risk of developing schizophrenia. A challenge in this small and until recent-generations isolated Oceanic population is that pedigrees are large and even those subjects who do not fall in the same close-knit pedigree are often found to be genetically related to some degree. Motivated by this observation we build a statistical test that can be applied to large pedigrees whose structure need not be known, one that evaluates the association between the extent of IBD sharing within specific genomic regions and the presence of schizophrenia. We first estimate a matrix of pairwise relatedness between subjects by measuring IBD sharing over the entire genome using Germline. This results in a matrix with genetic estimates of kinship amongst all pairs of subjects in the study, but the matrix is noisy. Applying a recently developed algorithm to the kinship matrix -- treelet covariance smoothing -- we reduce the noise by using the information inherent in clusters of individuals with estimated positive kinship coefficients. To form a test, we apply a kernel-based approach that compares local IBD sharing between cases and controls, while controlling for estimated relatedness. Within a given region, every pair of subjects is characterized by the segment-length inherited from a common ancestor (IBD). The statistic seeks to identify regions of the genome where the amount of IBD sharing amongst cases and/or amongst controls is greater than the sharing between cases and controls, because these regions have the potential to harbor mutations that are associated with schizophrenia. Our approach identifies regions of the genome as potentially harboring risk variants and some of those loci harbor known schizophrenia-associated genes, suggesting the results are likely to be meaningful. We continue to explore these data to identify variants affecting risk for schizophrenia.
Cumulative genetic load for known multiple sclerosis risk variants in Sardinia. A. Hadjixenofontos1, L. Foco2, P. A. Gourraud3, A. Tikka4, P. Bits5, R. Pastorino6, L. Bernardinelli7, L. J. L. McCauley1. 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Public Health, Neurosciences, Experimental and Forensic Medicine, Medical Statistics and Epidemiology Unit, University of Pavia27100 Pavia, Italy; 3) Department of Neurology, School of Medicine, University of California at San Francisco, San Francisco, CA, USA; 4) Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK; 5) Divisione di Neurologia, Ospedale S. Francesco, Nuoro, Italy; 6) Immunohaematology and Blood Transfusion Department, Ospedale S. Francesco, Nuoro, Italy.

Multiple Sclerosis (MS) is an autoimmune, demyelinating disease for which the prevalence rates follow a latitude gradient with the lowest prevalence in countries closest to the equator. Sardinia, in Insular Italy, is an exception to this trend with unenviable prevalence rates that range between 150 and 200 per 100,000 individuals. The geographical isolation of Sardinia presents a valuable opportunity to study a relatively homogeneous population, for which the underlying genetic load is not known. Recent collaborative investigations by the International Multiple Sclerosis Genetics Consortium (IMSGC) have increased the list of MS risk variants to 113 tag SNPs. We use these variants in the calculation of a MS genetic risk score with the goal of assessing the extent to which NE derived variants allow differentiation between Sardinian MS cases and controls. The genotyping data on our dataset of 16 Sardinian families and 94 unrelated Sardinian controls originated from a custom Illumina genotyping chip dubbed the ‘Immunochip’. Using Generalized Estimating Equations (GEE) to correct for correlations between relatives, we built a logistic regression model with disease status as the outcome, and the MS risk score, and sex as the predictors. An increase in the NE derived MS risk score was associated with MS case status in our preliminary analysis of 70 Sardinian cases and 244 Sardinian unaffected family members and population controls (OR=1.68, 95% CI=1.33-2.11, P<0.001). This is in contrast to our previous results with a version of the MS risk score that contrast to our previous results with a version of the MS risk score that included fewer tag SNPs and which was not associated with affection status. No differences in the MS risk score were detected between MS patients from Sardinian vs. Caucasian population (OR=1.11, 95% CI=0.81-1.58, P=0.55). In light of our latest results a more extensive analysis is currently underway. These will seek to capitalize on the relatedness in the sample by examining whether variants in the known regions are shared identical by descent (IBD) in cases from each family, and additionally will include information on rare variants from exome sequencing data which may be present in the known regions.

Interaction between Adiponectin and Adiponectin Receptor 1 is Associated with Age-related Hearing Impairment. C. Wu1,2, C. Tsai2, J. Hwang3, Y. Lu2,3, Y. Lin1,2, P. Chen1, W. Yang2,3, W. Liao2, Y. Lee2,3, T. Liu1, C. Hsu1, T. Liu. 1) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 4) Department of Otolaryngology, Buddhist Dalin Tzu-Chi General Hospital, Chiayi, Taiwan; 5) Institute of Biomedical Engineering, National Taiwan University, Taipei, Taiwan; 6) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 7) Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 8) Health Management Center, National Taiwan University Hospital, Taipei, Taiwan; 9) Research Center for Genes, Environment and Human Health, College of Public Health, National Taiwan University, Taipei, Taiwan.

Age-related hearing impairment (ARHI) is a complex disease caused by an interaction between environmental and genetic factors. Recently, several studies confirmed obesity as an independent risk factor for ARHI. In our previous study investigating the underlying mechanisms, we demonstrated that plasma adiponectin might protect peripheral hearing function (Clin Endocrinol (2011)). It has been revealed that polymorphisms of the adiponectin gene, ADIPOQ, might affect plasma adiponectin levels; and polymorphisms of both ADIPOQ and its type 1 receptor gene, ADIPOR1, have been related to obesity-related morbidities. Hence, we postulated that genotypes of ADIPOQ and ADIPOR1 might be associated with the development of ARHI. A total of 1682 volunteers (Han Chinese, aged 40 to 80 y) were included in the clinical analyses, and their audiological phenotypes were determined according to the Z scores converted from their original frequency-specific hearing thresholds. By using the database of Chinese Han haplotypes in International HapMap Project and NIEHS, followed by analyses with the Haploview software, 9 tagSNPs and 4 tagSNPs in ADIPOQ and ADIPOR1, respectively, were selected for genotyping. The genotypes were then correlated to the audiological phenotypes under the assumption of various inheritance models. Significant association was identified between certain ADIPOQ tagSNPs and Z scores under dominant, co-dominant, or additive models, whereas no association was identified between the ADIPOR1 tagSNPs and Z scores. Of note, the association between specific ADIPOQ tagSNPs and Z scores appeared to exist in subjects with specific ADIPOQ genotypes only, indicating a gene-gene interaction between ADIPOQ and ADIPOR1. These findings were then validated on haplotype analyses. Further measurement of plasma adiponectin level in 736 subjects revealed ADIPOQ genotypes might exert their effects on hearing levels through modulating the plasma adiponectin levels. In conclusion, the ADIPOQ genotypes were associated with ARHI, and might exert effects on hearing levels through modulating the plasma adiponectin levels. In contrast, although the ADIPOR1 genotypes were not directly associated with ARHI, the association between ADIPOQ and hearing thresholds were influenced by the ADIPOR1 genotypes. In other words, the development of ARHI might result from an interaction between adiponectin and type 1 adiponectin receptor.

The OPRM1 (opioid receptor, mu 1) gene has long been suspected to influence risk of heroin abuse. The 116A>G polymorphism, in particular, has been widely studied for its association with heroin and other addictions and its functional consequences, but the evidence for association is inconsistent. To identify other polymorphisms that may regulate OPRM1 and in turn influence risk of heroin abuse, we evaluated single nucleotide polymorphism (SNP) associations with OPRM1 expression to nominate cis-expression quantitative trait loci (cis-eQTL) in the human brain and subsequently tested the nominated SNPs for association with the heroin abuse phenotype. For the cis-eQTL analyses, we utilized SNP genotypes (Illumina Human1M-Duo or HumanHap660Y) and gene expression data from prefrontal cortex samples of normal human subjects in the publically available BrainCloud. Our cis-eQTL analyses used 112 African American and 109 Caucasian subjects from infancy through adulthood (ranging in age from <1 to 78 years old). In each ancestral group, 103 SNPs located within 100kb of OPRM1 were tested for association with OPRM1 expression level using linear models that accounted for sex and age within developmental stage (childhood vs. adulthood). Of the 103 SNPs, 14 were nominated as cis-eQTLs based on their association with OPRM1 expression (P<0.05) in either ancestral group. The 14 SNPs were then tested for association with heroin abuse using cases from the Urban Health Study and public controls (total N=7,095 African Americans and 3,824 Caucasians), who were genotyped on Illumina arrays and imputed using IMPUTE2 with reference to the 1000 Genomes ALL haplotype panel. In each ancestral group, SNP associations with heroin abuse were tested using logistic models accounting for sex and 3 ancestry-specific eigenvectors. Convergence of the SNP association results revealed that 4 of the nominated cis-eQTLs (P ranging from 0.015 to 0.046) were also associated with heroin abuse (P ranging from 6.17x10^{-5} to 3.85x10^{-3} in a meta-analysis combining the ancestry-specific results). The 4 newly identified regulatory SNPs had minor allele frequencies between 16% and 38% and consistent direction of association with heroin abuse across the ancestral groups. Our findings offer new insights into the role that OPRM1 may play in developing heroin abuse. Further studies are needed to confirm the observed SNP associations in other data sets and to elucidate their mechanisms.
988T Integrated eQTL and Disease-Associated eQTL Loci. P.J. Castaldi, M.H. Cho1,2, W. Qiu1, B.R. Celli2, J.H. Riley2, S.M. Fox2, D. Singh4, R. Tal-Singer4, B.A. Raby1, V.J. Carey1, E.K. Silverman1,2, C.P. Hersh1,2, ECLIPSE Investigators. 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 3) GlaxoSmithKline, Uxbridge, UK; 4) Medicines Evaluation Unit, University of Manchester, Manchester, UK; 5) GlaxoSmithKline, King of Prussia, PA, USA.

**Background:** Recent GWAS have identified novel loci associated with chronic obstructive pulmonary disease (COPD). Two of these loci have been shown to be associated with the expression level of nearby genes. We hypothesized that integrating eQTL and GWAS results from subjects with COPD would allow for the identification of novel COPD-associated loci, and that cis and trans eQTL analysis could characterize transcriptional networks of disease-relevant genes. **Methods:** Gene expression profiles from whole blood and induced sputum samples were generated for 121 subjects with COPD from the ECLIPSE study using the Affymetrix HG-U133 Plus 2.0 GeneChip. Genome-wide identification of cis eQTLs was performed using a cis window of 500kb. Multiple comparison adjustment was performed by controlling the false discovery rate at 10%. eQTL SNPs were tested for association with COPD susceptibility in a meta-analysis of 4 COPD case-control studies (NETT/NAS, ECLIPSE, GenKOLS, COPDGene). Targeted trans-eQTL analysis was performed for significant cis eQTL SNPs that were also associated with COPD. **Results:** Cis eQTL analysis in subjects with COPD identified 42,020 SNPs associated with expression of 7,223 unique genes in whole blood and 18,474 SNPs associated with 4,474 genes in sputum samples. Cis eQTL SNPs were significantly enriched in the top GWAS results (p<0.001 from both sources of expression data). When these SNPs were tested for association with COPD, SNPs associated with expression of 10 unique genes were also associated with COPD susceptibility (9 genes from blood, 2 genes from sputum with one gene present in both). All previously identified COPD GWAS loci harbored a significant cis eQTL. Trans eQTL analysis of these SNPs identified two SNPs associated with the expression of disease-relevant genes. Rs1828591 was associated in cis with HHIP (q value = 0.09) and in trans with RTN4R1 (q value = 0.03). Rs8034191 was associated with HAM and QTL analyses. Identification of additional PAH genes can be achieved by integrating these two analyses. **Conclusions:** Using a gene-based eQTL approach in whole blood and sputum, we identified ten SNP-gene functional units associated with COPD. Some previously described COPD loci are eQTL for more than one gene, and trans eQTL analysis identified that cis eQTL SNP for HHIP and IREB2 are also associated with regulation of genes in trans. Funding: K08HL102265, K08 HL097029, P01 HL105339, R01HL094635, R01NR013377, NCT00292552.
990W
Investigation of classical human leukocyte antigens (HLA) and expression quantitative trait loci (eQTL) within the HLA region in association with Parkinson’s disease. H. Payami,1,2, W. Wissemann,1 E. Hill-Burns3, C. Zabetian4, S. Factor2, N. Patsopoulos5,6,7, B. Boglund8, C. Holcombe9, G. Thomson9, H. Erlich8,10,1. 1 Division of Genetics, Wadsworth Center, New York State Department of Health, Albany, NY, USA; 2 Department of Biomedical Sciences, State University of New York at Albany, NY, USA; 3 VA Puget Sound Health Care System and Department of Neurology, University of Washington, Seattle, WA, USA; 4 Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA; 5 Program in Translational NeuroPsychiatric Genomics, Neurosciences Institute, Departments of Neurology, Brigham and Women’s Hospital, Boston, MA, USA; 6 Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA; 7 Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA; 8 Human Genetics Dept, Roche Molecular Systems, Pleasanton, CA, USA; 9 Department of Integrative Biology, University of California, Berkeley, CA, USA; 10 Childrens Hospital of Oakland Research Institute, Oakland, CA, USA.

Historically, tests of association between HLA and disease have been conducted with the ‘classical’ HLA alleles; i.e., polymorphisms that encode differential ability of HLA molecules to bind and present antigens. Association of PD with HLA, however, was discovered in a hypothesis-free GWAS and replicated using SNP-based studies. The original SNP that reached genome-wide significance, rs3129882, is in intron 1 of DRA and the other maps to intergenic sequences near DRA, DRB1 and DRB5. Our aim was to determine if these SNPs are tagging specific HLA alleles. We HLA-typed 2000 PD cases and 1898 controls on a standard panel of classical HLA alleles (PMID:20711177), and 843 cases and 856 controls (phs000126.v1.p1) for replication using two SNP-based imputation algorithms (de Bakker PMID:21051598 and HLA*IMP:02). We also sequenced 194 cases and 204 controls using 454-next-generation-sequencing, and calculated allele-specific sensitivity and specificity for each algorithm. For association studies we used the imputed data to increase sample size, mindful of allele-specific imputation accuracy. We tested association of HLA alleles and haplotypes with PD risk using conditional and stratified analyses to test interdependence of the classical variants identified here and the SNPs identified before. PD risk was associated positively with B*07:02, C*07:02, DRB5*01, B*11:01, and negatively with B*40:01, C*03:04, DRB4*01, DRB1*04:04, DAO1*03:01, DQB1*02:02 and alleles and haplotypes. However, when conditioned on the SNPs, the signals for the risk variants were abolished, but the ones for the protective variants remained nominally significant (Pconditioned<3E-10). With in-silico studies (SCAN, eQTLBrowser, Genevar), we found that PD-associated SNPs rs3129882 (original GWAS PMID:20711177), rs2395163 (US Meta analysis PMID:22451204) and rs660895 (French study PMID:22807207) are eQTLs for HLA-DR and HLA-DQ (4E:5:PoPeQTL:2e-79). In summary: Classical HLA-alleles do not explain the genetic variance is accounted for by single Major Histocompatibility Complex (MHC) locus, known as PSORS1. HLA-C is widely regarded as the strongest PSORS1 candidate gene, since markers tagging HLA-Cw*0602 consistently generate the most significant association signals in GWAS. However, the question as to whether HLA-Cw*0602 is the causal PSORS1 allele has not been resolved, especially as the role of SNPs that may affect its expression has not been investigated. Here, we have undertaken a comprehensive molecular characterization of the PSORS1 region, with the aim of identifying regulatory variants that may contribute to psoriasis susceptibility. We first analysed high-density SNP data and refined the PSORS1 locus to a 179kb region encompassing HLA-C and the neighbouring HCG27 pseudogene. Next, we compared multiple MHC sequences spanning the 179kb critical interval and catalogued 144 candidate susceptibility variants, which were unique to Cw*0602-bearing chromosomes. In parallel, we determined the epigenetic profile of the critical PSORS1 interval, uncovering three enhancer elements likely to be active in T lymphocytes. Finally, we showed that nine candidate susceptibility SNPs map within a HLA-C enhancer and that three co-localise with binding sites for immune-related transcription factors. This study suggests that the expression of HLA-Cw*0602 are likely to contribute to psoriasis susceptibility and highlight the importance of integrating multiple experimental approaches in the genetic analysis of complex genomic regions such as the MHC.

990T
Identification of candidate psoriasis susceptibility alleles within an HLA-C enhancer element. R.C. Trembath1,2, A. Clop2, A. Bertoni3, S.L. Span2, M.A. Simpson2, V. Pullabhata2, R. Tonda3, C. Hundraisen2, P. Di Meglio2, P. De Jong4, A. Hayday5, F.O. Nestle2, J.N. Barker5, R.J.A. Bell6, F. Capon7, 1 Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AD, United Kingdom; 2 Division of Genetics and Molecular Medicine, King’s College London, London SE1 9RT, UK; 3 Centre for Research in Agricultural Genomics (CRAG), Campus Universitat Autònoma de Barcelona, 08193 Cerdanyola del Valles, Spain; 4 BACPAC Resources Centre, Children’s Hospital Oakland, Oakland, CA 94609, USA; 5 Division of Immunology, Infection & Inflammatory Diseases, King’s College London, London SE1 9RT, UK; 6 Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA 94115, USA.

Psoriasis is a common inflammatory skin disorder that is inherited as a complex genetic trait. Although genome-wide association scans (GWAS) have uncovered more than 30 disease susceptibility regions, up to 50% of the genetic variance is accounted for by single Major Histocompatibility Complex (MHC) locus, known as PSORS1. HLA-C is widely regarded as the strongest PSORS1 candidate gene, since markers tagging HLA-Cw*0602 consistently generate the most significant association signals in GWAS. However, the question as to whether HLA-Cw*0602 is the causal PSORS1 allele has not been resolved, especially as the role of SNPs that may affect its expression has not been investigated. Here, we have undertaken a comprehensive molecular characterization of the PSORS1 region, with the aim of identifying regulatory variants that may contribute to psoriasis susceptibility. We first analysed high-density SNP data and refined the PSORS1 locus to a 179kb region encompassing HLA-C and the neighbouring HCG27 pseudogene. Next, we compared multiple MHC sequences spanning the 179kb critical interval and catalogued 144 candidate susceptibility variants, which were unique to Cw*0602-bearing chromosomes. In parallel, we determined the epigenetic profile of the critical PSORS1 interval, uncovering three enhancer elements likely to be active in T lymphocytes. Finally, we showed that nine candidate susceptibility SNPs map within a HLA-C enhancer and that three co-localise with binding sites for immune-related transcription factors. This study suggests that the expression of HLA-Cw*0602 are likely to contribute to psoriasis susceptibility and highlight the importance of integrating multiple experimental approaches in the genetic analysis of complex genomic regions such as the MHC.

992F
Joint eQTL Analysis in Multiple Tissues, Accounting for Heterogeneity and Incomplete Subject Overlap, G. Li1, A. A. Shabalin1, I. Rusyn2, F.A. Wright3, A.B. Nobel1,2. 1) Statistics and Operations Research, Univ North Carolina, Chapel Hill, NC, 2) University of North Carolina, Chapel Hill, NC, 3) University of North Carolina, Chapel Hill, NC, 4) Department of Biostatistics, University of North Carolina, Chapel Hill, NC.

Joint expression quantitative trait loci (eQTL) aims to identify genetic variants that regulate gene transcription, and help dissect complex transcriptional-based mechanisms of disease. Patterns of transcriptional variation are highly tissue-specific, and identifying common and specific eQTLs among various tissues is of great interest. However, most existing methodologies for multiple-tissue eQTL analysis are based on post-hoc analysis of eQTL from individual tissues, and therefore do not fully utilize the data structure of eQTL datasets consisting of multiple tissues. We propose a hierarchical model for the observed correlations of gene-SNP pairs across the available K tissues in a multi-tissue experiment, with a full specification of all 2^K possible patterns of null vs. alternative outcomes across the K tissues. The model allows for incomplete overlap of subjects across the tissues by fitting a flexible correlation structure for genotype-expression associations. The model is fit using an empirical Bayes approach, and provides interpretable posterior probabilities for eQTLs across the range of tissues, as well as control of the local FDR. Simulations demonstrate the approach is stable and computationally feasible for numerous tissues. We further illustrate using RNA-Seq expression data on 9 tissues and genotypes on 3.4 million SNPs, available in pilot form from the Genotyping-Tissue Expression project (GTEx) Consortium, with sample sizes per tissue ranging from 80-200.

Both inherited and acquired traits can be transmitted through multiple generations with some traits more stable than others. The relationship between the stability of such transgenerational inheritance and the genetic variations in an individual or cell has yet to be explored. We studied the effect of genetic polymorphisms on transgenerational inheritance of yeast segregants that were derived from a cross between a laboratory strain and a wild strain of Saccharomyces cerevisiae. For each of 2955 SNPs analyzed, the parental allele present in less than half of the 124 segregants panel was called the minor allele (MA). We found a nonrandom distribution of MAs in the segregants, indicating natural selection, as segregants with high MA content or amount (MAC) were not enriched with MAs from the parental strain that contributed significantly more to the whole set of MAs. We compared segregates with high MAC relative to those with less and found a more dramatic shortening of the lag phase length for the high MAC group in response to 14 days of ethanol training. Also, the short lag phase as acquired and epigenetically memorized by ethanol training was more dramatically lost after 7 days of recovery in ethanol free medium for the high MAC group. Sodium chloride treatment produced similar observations. Using public data, we found more complex linkage to mRNA expression of hundreds of genes. Finally, by analyzing a recently published dataset of 1009 yeast segregants that identified numerous additive QTLs for 46 traits, we found by multivariate regression analysis preferential MAC linkage to traits with high number of known additive QTLs (average 16 for the 5 MAC-linked traits vs. average 2 for traits without a MAC QTL). The contribution of traits via the additive effects of a large number of minor effect SNPs and may thereby account for the 'missing heritability' in most complex traits. These results provide evidence for the slightly deleterious nature of most MAs and a lower capacity to maintain inheritance of traits in individuals or cells with greater MAC, which have implications for disease prevention and treatment. Individuals with high MAC may be more susceptible to environmental pathogens, but they may also be more treatable if treatment was administered relatively early before the disease has progressed past a threshold of no return, because the acquired disease trait may be less stably maintained in these individuals.

995F High-dimensional genetic prediction of type-2 diabetes susceptibility. Y. Klimentidou1, A. Vazquez2, G. de los Campos3, D. Allison2. 1) Division of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ; 2) Section on Genetic Epidemiology, Human Genetics Institute, University of California Los Angeles, Santa Monica, CA; 3) Department of Statistical Genetics, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL; 3) Office of Energetics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL.

Recent findings from genome-wide association studies (GWAS) have uncovered approximately 65 loci confidently associated with type-2 diabetes (T2D) risk. However, these loci explain less than ten percent of the expected heritability of T2D, and have limited predictive value. Different statistical methods along with the inclusion of additional variants may increase the utility of genomic information, and ultimately lead to a more confident identification of individuals at highest genetic risk of T2D. In this study, we use genotype information from up to approximately 2.5 million SNPs (single nucleotide polymorphisms), along with information about each SNP's potential importance in T2D risk, as determined by the summary statistics of a recent meta-analysis of GWAS. We build and test various polygenic prediction models in the context of early-life prediction of T2D susceptibility using 1) G-BLUP (genomic best linear unbiased predictor) and 2) multi-SNP risk scores. Prediction models are developed and tested within the Framingham Heart Study dataset, and the main results are replicated in the ARIC (Atherosclerosis Risk in Communities) dataset. We find that prediction accuracy increases substantially with the inclusion of additional SNPs, from an AUC (area under the receiver operating characteristic curve) of 0.689 for the commonly used 'top hits' genetic risk score, up to 0.814 using a multi-SNP risk score that includes over 3,700 SNPs. Our findings are successfully replicated in the ARIC dataset. We have shown that T2D genetic risk prediction models can be markedly improved with the inclusion of thousands of SNPs, given prior information about relative SNP importance.

994T Genetic association studies have successfully identified thousands of loci in hundreds of phenotypes. In order to increase power studies typically utilize one degree of freedom statistical tests, making a restrictive additive (or multiplicative) assumption about the underlying genetic basis of the phenotype of interest. In this work we examine the validity of this choice, introducing new methods to determine how much better a given phenotype is modeled with unrestricted effect sizes for zero, one, or two copies of each causal variant. We know from both human and model organism genetics that many variants behave in a dominant, recessive, or over-dominant manner and these rich heritability models of disease affect complex phenotypes is currently unknown. Our approach is an extension of the recent mixed linear model approaches to estimating the phenotypic variance driven by genotyped variants (Yang et al Nat Gen 2010), which also assume a one degree of freedom model. We propose multi-variability component methods, that estimate the additional phenotypic variance captured by increasingly freer disease models. The variance components are constructed by measuring the genome-wide similarity of individuals according to different encodings of genotypes. The method has a corresponding likelihood based test with a null model of additivity, allowing a formal examination of the additive assumption. We show via extensive simulation over real and simulated data that our test statistic has the appropriate type-I error rate and actualizes the fraction on phenotypic variance driven by each disease model. We apply our method to phenotypes from the National Finnish Birth Cohort and WTCCC data sets, show several potential sources for bias including ascertainment data sets, relatedness, and population structure, and confirm the known dominance structure in Type 2 Diabetes (p-value < 5e-5). We derive a BLUP based phenotypic prediction method from our mixed linear model, which will improve our prediction ability over existing additive models for those phenotypes with more complex underlying structure. This work will add to our fundamental understanding of the genetic basis of complex human phenotypes and offer insights into the missing heritability problem.

996W Investigating missing heritability and improving risk prediction with maximally free marginal models of phenotype. N. Zaitlen1, T. Wingo4, A. Gusev1, D.J. Cutler2. 1) Department of Medicine, University of California San Francisco, San Francisco, CA; 2) Department of Neurology, Emory University School of Medicine, Atlanta, GA; 3) Department of Genetic Epidemiology, Harvard School of Public Health, Boston, MA; 4) Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America.

Genetic association studies have successfully identified thousands of loci in hundreds of phenotypes. In order to increase power studies typically utilize one degree of freedom statistical tests, making a restrictive additive (or multiplicative) assumption about the underlying genetic basis of the phenotype of interest. In this work we examine the validity of this choice, introducing new methods to determine how much better a given phenotype is modeled with unrestricted effect sizes for zero, one, or two copies of each causal variant. We know from both human and model organism genetics that many variants behave in a dominant, recessive, or over-dominant manner and these rich heritability models of disease effect complex phenotypes is currently unknown. Our approach is an extension of the recent mixed linear model approaches to estimating the phenotypic variance driven by genotyped variants (Yang et al Nat Gen 2010), which also assume a one degree of freedom model. We propose multi-variability component methods, that estimate the additional phenotypic variance captured by increasingly freer disease models. The variance components are constructed by measuring the genome-wide similarity of individuals according to different encodings of genotypes. The method has a corresponding likelihood based test with a null model of additivity, allowing a formal examination of the additive assumption. We show via extensive simulation over real and simulated data that our test statistic has the appropriate type-I error rate and actualizes the fraction on phenotypic variance driven by each disease model. We apply our method to phenotypes from the National Finnish Birth Cohort and WTCCC data sets, show several potential sources for bias including ascertainment data sets, relatedness, and population structure, and confirm the known dominance structure in Type 2 Diabetes (p-value < 5e-5). We derive a BLUP based phenotypic prediction method from our mixed linear model, which will improve our prediction ability over existing additive models for those phenotypes with more complex underlying structure. This work will add to our fundamental understanding of the genetic basis of complex human phenotypes and offer insights into the missing heritability problem.
Life course variations in the heritability of body size. J. Zhao, J.A. Luhan, S.J. Sharp, R. Hardy, A. Wong, Q. Tan, N.J. Wareham, D. Kuh, K.K. Ong. 1) MRC Epidemiology Unit, University of Cambridge, Institute of Metabolic Science, Box 285, Addenbrooke’s Hospital, Hills Road, Cambridge CB2 0QQ, UK; 2) 2MRC Unit for Lifelong Health and Ageing, University College London, London WC1B 5JJ, UK; 3) Epidemiology and Biostatistics, University of Southern Denmark, J.B. Winsloes Vej 9, DK-5000 Odense C, Denmark.

Background: It has been shown recently that whole genome data can facilitate estimation of genetic contributions to a variety of traits via a mixed model framework as implemented in GCTA and R/SAS (Yang et al. Nat Genet 2010, 42:565-9; Zhao & Luan. J Prob Stat 2012. doi 10.1155/2012.485174). Our aim was to use this approach to investigate the life course variations in heritability of body size. Methods: We analysed height, weight and body mass index variables at 11 time-points in 2,452 individuals (1,225 men, 1,227 women) born in 1946 and enrolled in the MRC National Survey of Health and Development (NSHD), with genotypes at 147,949 single nucleotide polymorphisms (SNPs) on Metabochips which were subsequently imputed to 506,255 according to the 1000Genomes project. We obtained genome-wide kinship matrices using genotypes at SNPs on Metabochips and genotypes at all SNPs, which were used in mixed models as implemented in the computer program GCTA. Results were also compared to those obtained using an alternative procedure of kinship estimation in PLINK and mixed models in R. Results: In line with earlier findings that specific genetic variants have variable temporal effects in this cohort (Hardy et al. Hum Mol Genet. 2010; 19:545-552), we observed age-related variations in heritability estimates. Estimates based on genotypes at SNPs on Metabochips and genotypes at all SNPs were comparable but generally lower than recently reported GCTA estimates with mean(range) being 0.09(0-0.50), 0.11(0-0.20), 0.10(0-0.22) for height, weight and body mass index, respectively. Variation in estimates was also seen between alternative procedures. Conclusion: This work supports the utility of large-scale genotypes in heritability estimation and highlights the age-related variability in genetic contributions to body size across the life course. Further work will be to distinguish the effects of established variants and to consider the role of rare variants in a unified longitudinal model including contrast with models assuming various degrees of temporal homoscedasticity.
Assessing the Impact of Coding Variants on Lipid Levels with the Exomechip. K.E. Stirrup 1, N. Nasca 2, D. Pasko 1, A. Mahajan 3, S. Kanoni 3, UK Exomechip Lipid Consortium. 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 2) Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; 3) NHRI Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK; 4) Genetic and Molecular Epidemiology Unit, Lund University Diabetes Center, Malmo, Sweden; 5) MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK; 6) School of Public Health, Imperial College London, London, UK; 7) Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK; 8) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

Multiple rounds of Genome Wide Association (GWA) meta-analyses have so far implicated mainly common variants underlying the molecular basis of blood lipid levels in humans, explaining only a fraction of the genetic component of these traits. The aim of the Illumina HumanExome chip (247,870 SNPs) was to provide a cost-effective way to assess the genetic contribution of low frequency (minor allele frequency (MAF) < 5%) and rare (MAF < 1%) exonic variants which are likely to be functional. An estimated 97% of non-synonymous variation and 94% of splice and stop variation detected in an average genome through exome sequencing are covered by the design. The array also includes a scaffold of common tag SNPs, HLA and mitochondrial markers as well as all known GWAS lead SNPs (NHGRI catalogue).

We have set up a collaborative effort to perform an inverse variance weighted meta-analysis in European and South-East Asian cohorts (current analysis includes ~20,000 samples) for high-density lipoprotein cholesterol (LDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides (TG) and total cholesterol (TC) levels. Genotype calling was performed using GenCall followed by zCall. Meta-analysis (autosomal data) was performed in GWAMA, testing fixed and random effects. We excluded subjects on lipid lowering medication and analysed disease and control cohorts separately.

We implemented additional quality control steps for filtering genotypes obtained for rare variants. In total, ~175,000 variants passed QC criteria. We then examined SNPs which were associated with lipid traits at a significance threshold (p < 5 × 10−8) and LD of the sentinel SNP, these are good functional candidates. Interestingly, we found a lipid trait other than the one analysed. We identified a coding variant with 2 known loci showing a stronger association than the reported sentinel SNP, these are good functional candidates. Interestingly, for 21 known loci we found a variant which was not in LD with the sentinel SNP and remained significant upon conditional analysis, most of them (76%) were low frequency / rare variants. We are currently pursuing replication approaches for 74 (78%) are coding SNPs putative new signals of which ~64% have MAF <1%.

Narrowing the gap on heritability of common disease by direct estimation in case-control GWAS. D. Golan, S. Rosset. Department of Statistics, Tel-Aviv University, Tel-Aviv, Israel, 69978.

One of the major developments in recent years in the search for missing heritability of human phenotypes is the adoption of linear mixed-effects models (LMMs) to estimate heritability due to genetic variants which are not significantly associated with the phenotype. A variant of the LMM approach has been adapted to case-control studies by Lee et al. (2011) and applied to many major diseases, successfully accounting for a considerable portion of the missing heritability. For example, for Crohn’s disease the estimated heritability was 42.5%, considerably higher than the heritability accounted for by significantly associated SNPs. We propose to estimate heritability of disease directly by regression of phenotype similarities on genotype correlations, corrected to account for ascertainment (the enrichment of cases in the study compared to the population). We refer to this method as genetic correlation regression (GCR). Using GCR we estimate the heritability of Crohn’s disease at 34% using the same data as Lee et al., much closer to the estimated heritability of 56-60% obtained from family studies. GCR estimates of heritability are also considerably higher than LMM for bipolar disorder (54% compared to 43%, respectively). To understand the gap between GCR and LMM estimates, we conduct extensive simulations to evaluate the performance of both heritability estimation methods. Contrary to previous simulations conducted in the context of heritability estimation, our simulations are the first to simulate the entire generative process of data collection, including generating genotypes and selecting cases and controls. This realistic simulation scheme captures some essential features of case-control GWASs which were overlooked by previous simulation schemes. For example, while the true heritability is rare (MAF < 1%), the heritability is rare (MAF < 1%) for randomly sampled cases tend to be more genetically similar than random controls, even when they are unrelated. Our simulations demonstrate that GCR produces accurate and unbiased estimates of heritability for a wide range of scenarios, while LMM estimates are considerably and consistently negatively biased under ascertainment. We thus conclude that GCR estimates of heritability are more reliable than LMM estimates. Last, we develop a heuristic correction to LMM estimates, which can be applied to published LMM results. Applying our heuristic correction increases the estimated heritability of multiple sclerosis from 30% (Lee et al. 2013) to 52.6%.

GWAS-data based heritability estimation of memory related phenotypes. C. Vogler1,2,4, V. Freytag2, A. Milnik2, L. Gschwind2, D. Coy넬3, A. Heck1,2, D. de Quervain1,2, A. Papassotropoulos1,2,4, 1) Psychiatric University Clinic, University of Basel, Basel, Basel City, Switzerland; 2) Molecular Neuroscience, University of Basel, Basel, Basel City, Switzerland; 3) Cognitive Neuroscience, University of Basel, Basel, Basel City, Switzerland; 4) Department Biocentrum, Life Sciences Training Facility, University of Basel, Basel, Switzerland.

Background: Genome-wide association studies provide the means to identify the molecular underpinnings of complex heritable traits. Heritability is a concept that summarizes how much of the phenotypic variation in a trait is attributable to genetic factors. Conventionally, heritability estimates in humans are derived from phenotypic data solely by computing correlations between relatives. However, those methods imply that the amount of resemblance that is due to shared environment is the same for different degrees of relationship. Further, the genetic identity between sib pairs is assumed to be 50%, whereas it has shown to vary from ~40% to ~60% (Visscher, 2006). Alternatively, recently developed methods, propose to infer genetic identity from high-throughput SNP data and to correlate these estimates with phenotypic resemblance among unrelated individuals. This allows heritability estimation for any specific GWAS dataset. Here we present genetic marker based heritability estimates for working and episodic memory phenotypes.

Methods: A healthy young Swiss sample comprising a total of N=1789 individuals was subjected to phenotypic assessment with a cognitive testing battery and subsequent genotyping using the Affymetrix SNP Array 6.0. Working memory (WM) performance was measured with the working memory (WM) performance was measured with the n-back paradigm. Episodic memory was assessed with the International Affective Picture System (IAPS) and lists of semantically unrelated words taken from the collections of Hager and Hasselhorn (Hager and Hasselhorn, 1994). In order to obtain the heritability estimates, we performed restricted maximum likelihood (REML) analyses implemented in the GCTA software package (Yang et al. 2011).

Results: Heritability estimates occupy a range from 26 to 65% with a mean standard error of 19%. The average heritability estimate for the attention corrected WM performance is 60% (SE: 19%) and the mean heritability for the different episodic memory tasks is 49% (mean SE: 20%). These estimates are consistent with previously reported heritability for WM (Ando, 2001) and episodic memory (Papassotropoulos, 2011). As this reflects that heritability for the cognitive phenotypes is represented in the GWAS data, this design can serve to unravel the molecular players that act in concert to form these complex traits.
1002W
Polygenic analysis of type II diabetes in the DIAGRAM3 GWAS meta-analysis, and partitioning of polygenic signal in tissue-specifically active genes.

A.I. Doney1, J. Wang1, B.F. Voight1, E. M. Kane1, P. F. Sullivan1, S. Roussos1,2, O. Gottsmann2, A. Morris3, R. Loos1,3, E. Bottlinger1, S. Raychaudhury3, M. McCarthy3, E.A. Stahl3,6,7, the DIAGRAM Consortium. 1) Institute for Genomics and Multiscale Biology, Department of Genetics and Genomics, School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 6) Partners HealthCare Center for Personalized Genetic Medicine, Brigham & Women’s Hospital, Harvard Medical School, Boston, MA; 7) Broad Institute, Cambridge, MA; 8) Center for Statistical Genetics, Icahn School of Medicine at Mount Sinai, New York, NY; 9) Dept. of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

INTRODUCTION: Polygenic risk score analyses have shown that a substantial portion of the missing heritability of common, complex diseases can be explained by weakly associated GWAS SNPs that fail to reach genome-wide significance. Separate analyses have demonstrated that tissue specific chromatin marks can be used to identify causal tissues by enrichment in GWAS associations (Trynka 2013). We combine these approaches, partitioning polygenic scores across liver- and pancreatic islet-specific gene sets in order to dissect the genetic architecture of type 2 diabetes (T2D), a common metabolic disorder that affects up to 8% of the adult population.

METHODS: We first conducted polygenic risk score analysis of T2D using the DIAGRAM GWAS (Morris 2012; 12,171,567 European cases/controls). Polygenic risk scores based on LD-pruned SNPs stratified by discovery GWAS P and risk allele frequency (RAF) were tested in either WTCCC1 (1,924/2,938 cases/controls) or the Mount Sinai Biobank (4,095/9,832 multiplex P=0.5), respectively, each corresponding to 3.3% of LD-pruned SNPs (transcripts +/- 100Kb). Top polygenic scores for these SNP sets are significant (liver P=0.002, R=0.003; pancreatic P<0.001, and with many additional scores P<0.05). We then identified MACS-called H3K4me3 ChIP-seq peaks from the NIH Epigenomics Roadmap project (34 tissues) in the promoter regions of Refseq transcripts (TSS +/- 1Kb), and normalized to obtain tissue-specificogenicity scores. We have explored distance from transcription start site as well as the nearest recombination hotspots, to define SNP scripts, to identify tissue-specific gene sets for partitioned polygenic analyses. RESULTS: Polygenic scores were significantly associated with T2D in the test data, with the strongest score having P=3×10−25 and R=0.035 (841 SNPs with PGWAS<0.001, and with many additional scores P<0.05). Moreover, scores with PGWAS<0.001 and RAF<0.88. 761 (1.80%) and 442 transcripts (1.05% of all transcripts) constitute primary liver and pancreatic islet gene sets (specificity score >0.5), respectively, each corresponding to 3.5% of LD-pruned SNPs (transcripts +/-100Ko). Top polygenic scores for these SNP sets are significant (liver P=0.002, R=0.002, 28 SNPs with PGWAS<10−23; pancreatic P<0.001, R=0.001, 817 SNPs with PGWAS<0.05), and show 1.5- to 2.1-fold greater R2 than expected for the proportion of SNPs represented. CONCLUSIONS: Tissue-specific compartments of polygenic signal can be derived from gene harboring chromatin marks, and will be applied in further studies of T2D causal mechanisms including the relative roles of liver versus pancreatic islets, and pleiotropy with T1D and cardiac/metabolic phenotypes.

1004F
The genetic architecture of schizophrenia in the Swedish Schizophrenia Study.

E.A. Stahl2,3, S.H. Lee2, G. Moser2, S. Ripke2,3, D.M. Rudfer2,4, J. Lee5, J. E. T. Law2, E.A. Stahl2,3, S. Mallar3, E. Bottinger1, S. Raychaudhury2, M. McCarthy3, O. Gottsmann2, A. Morris3, R. Loos3, E. Bottlinger1, S. Raychaudhury2, M. McCarthy3, E.A. Stahl3,6,7, the DIAGRAM Consortium. 1) Department of Psychiatric Genomics in the Department of Psychiatry, Friedman Brain Institute, and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, NY, NY 10029; 2) The University of Queensland, Queensland Brain Institute, Brisbane, QLD, 4072, Queensland, Australia; 3) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, 02114; 4) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, 02142; 5) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 6) Department of Genetics, Psychiatry, and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599.

INTRODUCTION: Schizophrenia has long been recognized to be familial, and more recently to be highly polygenic. The substantial contribution to phenotypic variance of SNPs falling short of genome-wide significance in GWAS is well appreciated, particularly in schizophrenia. Here we describe inference analyses of schizophrenia genetic architecture using the comprehensive genomic data collected on Swedish national registry samples.

METHODS: We analyzed GWAS data of 5,001 schizophrenia cases and 6,243 healthy controls from the Sweden (imputed to 1000 Genomes, 5.7M SNPs after QC) as test data in polygenic risk score (PRS) analyses with PGC schizophrenia GWAS meta-analysis results for training data. Approximate Bayesian Polygenic Analysis (ABPA) was performed under a mixture model of independent associated and null SNPs (Stahl 2012 PMID 22446960). Linear mixed models of normally distributed effects were fit using GCTA (Yang 2011 PMID 21167468, Lee 2011 PMID 2176301), and LD-based S.M. Purcell1,2, P. Sklar1,4, E. A. Stahl2,3, O.Gottesman1,3, E. Bottinger1, S. Raychaudhury2, M. McCarthy3, E.A. Stahl3,6,7, the DIAGRAM Consortium. 1) Human Genetics, Institute of Clinical Genetics, School of Medicine, University of Southern Denmark and Odense University Hospital; 2) Department of Medicine/Division of Cardiology, David Geffen School of Medicine, UCLA, LOS ANGELES, CA; 3) Department of Human Genetics, David Geffen School of Medicine, UCLA, LOS ANGELES, CA; 4) Department of Chemistry and Biochemistry, University of Colorado Boulder, CO; 5) Howard Hughes Medical Institute, Boulder, CO.

Over the last few years, the impact of the gut Microbiome composition in complex traits has received a lot of attention. The gut Microbiome hosts a multitude of bacteria species with a tremendous genetic diversity and has shown to regulate complex traits in addition to the host genome. Using the Hybrid Mouse Diversity Panel (HMMP) that consists of more than 100 inbred strains of mice, we show that the gut Microbiome composition explains a considerable portion of the variation in metabolic traits. We further partition the heritability of metabolic traits to the contribution of genomic regions and show that these estimates may be different if we account for the gut Microbiome. Our findings suggest that gut Microbiome may be one of the key factors in the missing heritability of complex diseases.
1006T
Proteomic and bioinformatics analysis of the endolysosomal pathway: Deciphering the insulin receptor Golgi/endosome subnetwork (IRGEN). M.B. DJIOU1, C. LANDRY2, R. Laframboise3, R. Faure1. 1) PEDIATRICS, CHUQ Research Center CRCCHUQ (Centre-Mère-Enfant-Soleil)/University Laval, Quebec, Quebec, Canada; 2) Institut de Biologie Intégrative des Systèmes (IBIS), Université Laval, Quebec, PQ, Canada.

While insulin resistance can be a heritable trait in lean and obese subjects, it remains unclear which cellular mechanism is the dominant driver that determines insulin sensitivity under baseline conditions. The study of the proteomes of lower organisms recently confirmed that proteins are not randomly connected; most interact with just a few others, whereas a limited number (hubs) interact with many others. This nodular architecture is conserved and provides a robust defense against random alterations, but connected proteins-which tend to play essential roles in organisms-can also be identified in candidate gene studies and functionally assessed in lower organisms recently confirmed that proteins are not randomly connected; most interact with just a few others, whereas a limited number (hubs) interact with many others. This nodular architecture is conserved and provides a robust defense against random alterations, but connected proteins-which tend to play essential roles in organisms-can also be identified in candidate gene studies and functionally assessed in lower organisms.
Whole genome analysis in fibromyalgia suggests a role for the central nervous system in disease susceptibility. E. Docampo1,2,3, G. Escar- 
amis1,2,3, M. Gratacos1,2,3, S. Villafloros1,2,3, P. Puig1,2,3, M. Kogevinas1,3, A. Collado1, J. Carbonell1, R. Rivera2, J. Vida2, A. Alegre6, R. Rabionet1,2,3, X. Estivill1,2,3 | 1) Genetic Causes of Disease Group, Centre for Genomic Regulation (CRG), Barcelona, Spain; 2) Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) Centre for Translation and Personalized Medicine in Red Epidermo- logia y Salud Publica (CIBERESP), Barcelona, Spain; 4) CREAL, Centre de Recerca en Epidemiologia Ambiental, CREAL, Barcelona, Spain; 5) Rheumatology Service, Hospital Clinic, Barcelona, Spain; 6) Rheumatology Unit, Instituto Provincial de Rehabilitacion, Madrid, Spain; 7) Rheumatology Unit, Rheumatology, Guadalajara, Spain; 8) Chronic Fatigue Syndrome Unit, Internal Medicine, Barcelona, Spain.

Fibromyalgia (FM) is a highly disabling syndrome defined by a low pain threshold and a permanent state of pain. The mechanisms explaining this complex disorder remain unclear and its genetic factors have not been identified. The aim of this study was to elucidate genetic susceptibility factors for FM. We used the Illumina 1 million duo array to perform a genomewide association study (GWAS), and Affymetrix's 2X400K platform for array compar- ative genomic hybridization (aCGH) to identify regions varying in copy number that could be involved in FM susceptibility. GWAS was performed in 300 FM cases and 203 controls. No SNP reached GWAS significance, but 21 of the most associated SNPs were chosen for replication in over 900 cases and 900 pain free-controls. Four of the strongest associated SNPs selected for replication showed a nominal association in the joint analysis, and one, rs11127292 (MYT1L) was found to be associated to FM with low comorbidities (p=2.8x10−5, OR (95%CI)=0.58 (0.44-0.75). By aCGH, an intronic deletion in NRXN3 showed to be associated to female cases of FM with low levels of comorbidities (p=0.021, OR (95%CI)=1.46 (1.05-2.04)). Both GWAS and aCGH results point at a role for the central nervous system in FM pathogenesis. The identification of genes whose expression changes could be involved in the core features of FM will contribute to a better understanding of this disorder, currently considered musculoskeletal and affective. This work was supported by ENGAGE, NOVADIS (EU and Spanish Ministry of Economy and Competitiveness).

Sex-specific heritability of traits related to human obesity in rhesus macaques. A. Vinson1,2, A.D. Mitchell1, D. Toffey3, M.J. Raboin1 | 1) Molecular and Medical Genetics, Oregon Hsc & Sci Univ, Portland, OR; 2) Oregon National Primate Research Center, Oregon Hsc & Sci Univ, Portland, OR; 3) Division of Genetics, Oregon National Primate Research Center, Oregon Hsc & Sci Univ, Portland, OR.

Obesity is a major component of the metabolic syndrome and is associated with insulin resistance, type 2 diabetes mellitus, and increased risk for cardio- vascular disease. Numerous studies in humans have demonstrated a genetic contribution to obesity. Further, obesity is sexually dimorphic in humans, and sex-specific genetic effects on human obesity have also been reported. The goals of this study were to explore the utility of the rhesus macaque as a genetic model for human obesity, by assessing the extent of additive genetic effects on traits related to human obesity and the influence of sex on these genetic effects. We measured weight, body mass index (BMI), abdominal circumference, and waist-to-thigh ratio in 584 Indian-origin females, P=0.07), abdominal circumference (h2=0.273 in females, P=0.03), vs. h2= 0.198 in males, P=0.07), abdominal circumference (h2=0.258 in females, P=0.05, vs. h2= 0.101 in males, P=0.25), and waist-to-thigh ratio (h2= 0.273 in females, P=0.03, vs. h2= 0.191 in males, P=0.06). We conclude that genetic contributions to traits related to human obesity can be detected and mea- sured in rhesus macaques, and that genetic contributions to some traits may differ by sex. Our results are consistent with substantial similarity between macaques and humans in the genetic architecture of obesity, and support the value of the rhesus macaque as a genetic model for this human condition.

Gene expression changes in response to paraquat and caffeine in a Drosophila model of Parkinson’s Disease. E. M. Hill-Burns1, P. Ganguly4, W.T. Wissmann1, W.J. Wolfgang5, H. Payami1,2,3 | 1) Division of Genetics, NY State Department of Health Wadsworth Center, Albany, NY; 2) Division of Infectious Disease, NY State Department of Health Wadsworth Center, Albany, NY; 3) Department of Biomedical Sciences, State University of New York Albany, NY.

Previous studies have established that paraquat exposure can induce Parkinsonism in Drosophila and serve as a model for Parkinson’s disease (PD) in humans. Consistent with findings in humans that coffee drinking reduces risk of PD with a 25-30% lower risk of PD (in males) and a 60-70% lower risk of PD (in females), our data support a protective role of caffeine in the rhesus paraquat model, observing that exposure to caffeine can partially rescue paraquat-induced mortality. To better understand the mechanisms at work in these Drosophila models and how they might relate to PD pathology in humans, we sought to identify genes whose expression changes significantly in response to paraquat and/or caffeine. Flies were kept on food containing either no caffeine or 1mM caffeine for ten days, and then transferred to food containing either no paraquat or 5mM paraquat (in addition to the original concentration of caffeine), for a total of four treatment conditions with 3-4 replicates of each of 30 female flies per vial. Fly heads were dissected after six days of paraquat treatment (or no paraquat treatment), RNA was extracted from the pooled tissue of each replicate, and cDNA was synthe- sized and hybridized onto arrays. Expression of 18,954 transcripts, using Affymetrix expression arrays was assayed in dissected heads of untreated flies, flies exposed to paraquat, flies exposed to caffeine, and flies exposed to both paraquat and caffeine. Array data were corrected for background signal and normalized. We tested for differences in expression associating with sex, with caffeine (untreated vs. caffeine-only-treated flies), or for paraquat-caffeine interaction in all four treatment conditions. We found transcripts of two genes with significantly different expression in response to paraquat exposure: CG3016 (Punadjusted=4E-6 and Punadjusted=0.04) and CG3016 (Punadjusted=4E-6 and Punadjusted=0.04). We also found transcripts of two genes with significantly different expression upon caffeine exposure: Hr38 (Punadjusted=3E-6; Padj- adjusted=0.04) and CG3016 (Punadjusted=4E-6; Padjadjusted=0.04). No transcripts showed significant evidence for paraquat-caffeine interaction on expression when adjusted for genome-wide tests. Human orthologs for these Drosophila genes include N4a2 (with Hr38) and USP30 (ubiquitin specific peptidase 30, with CG3016). Both N4a2 (aka Nurr1) and ubiquitin-related genes have been previously linked to PD in humans, making these genes interesting candidates for further study.

Probing the Gaucher/Parkinson Link By Crossing Different Mouse Models. N. Tayebi1, E. Maniwang1, R. Tamargo1, Y. Blech-Hermoni1, N. Moaven1, E. Allahi1, B. Berhe2, E. Sidrinsky1 | 1) Molecular Neuroge- netics, Medical Genetics Branch, NHGRI, NIH, Bethesda, MD; 2) Department of Biomedical Sciences, State University of New York at Stony Brook, Stony Brook, NY.

Gaucher disease, the most common lysosomal storage disorder, results from deficient glucocerebrosidase (GBA1) causing pathologic accumulation of glucosylceramide and glucosylsphingosine. Mutations in GBA1 are identified in an increased frequency in patients with Parkinson disease (PD) and Dementia with Lewy Bodies (DLB), although the mecha- nism by which they are associated is not fully understood. Three novel GBA1 mutations identified in subjects with PD are the loss-of function muta- tions, c.84dupG and IVS2+1. To explore how loss-of-function mutations on GBA1 alters GlcCer levels and enzyme activity were determined. In a second set of animals with the same copy number in the -synuclein (a-syn) copy number in the TgA53T-syn mice were used as controls. The -syn copy number in the TgA53T-syn/wt and TgA53T-syn mice (a-syn) copy number in the TgA53T-syn mice (a-syn) copy number in the TgA53T-syn mice, which were crossed into the GBA1−/− background. One allele might impact PD pathogenesis, we attempted to model this situa- tion in mice using null allele gba carrier knock-out mice (gba−/−/wt) crossed to wild-type a-syn mice. Two transgenic genotypes were confirmed, and nine sets of mice were fol- lowed over two years, monitoring for weight and neurological symptoms. Weight loss was a common initial sign, followed by arching back, impaired axial rotation, and finally lack of any movement. Mice with progressive neurologic deterioration were euthanized with their respective controls, and midbrain, dorsal cortex, and spinal cord were collected. GBA expression, glucocerebrosidase (GCase) levels and enzyme activity were determined. Over-expression of a-syn was confirmed by western blot. Levels of RNA expression and GCase protein expression in the brain samples were not significantly different upon caffeine exposure: Hr38 for both a-syn copy number in the TgA53T-syn mice (a-syn) copy number in the TgA53T-syn mice, which were crossed into the GBA1−/− background. One allele might impact PD pathogenesis, we attempted to model this situa-
**1017W**

Cell-specific enrichment metrics for overlap of signals from GWAS with DNase hypersensitivity sites. D. Chasman, F. Giuliani on behalf of the CKDGen and BP Metabochip-ICBP consortia. Division of Preventive Medicine, Brigham & Women’s Hospital, Boston, MA.

SNP associations from GWAS are enriched in non-coding regions mapping to chromatin marks. Quantitative characterization of this enrichment by cell type may identify causal SNPs and help reveal the physiologic basis of a phenotype of interest. We propose quantifying enrichment as an odds-ratio (OR) comparing the odds of a phenotype’s GWAS SNPs overlapping the chromatin marks of a particular cell with a reference odds. The reference odds may pertain either to reference SNPs mapping to chromatin marks of the same cell or to the same phenotype’s SNPs mapping to chromatin marks of a reference cell. We also considered a related notion of specificity quantified by relative entropy (H), the amount of information in bits needed to distinguish the enrichment in chromatin marks of phenotype SNPs compared with reference SNPs. We explored these metrics in publicly available data from 123 cell types, in which 1.1-6.7% of autosomal HapMap SNPs used by GWAS studies map to DNase hypersensitivity sites (DHSs). Applied to the NHGRI GWAS catalog, the enrichment measures were consistent with known biology. For example, across all 123 cell types and using reference SNPs from the GWAS catalog excluding the phenotype of interest, breast cancer SNPs were most enriched in DHSs from mammary ductal carcinoma epithelial cells (62% SNPs, OR= 4.4, p=0.012, H=0.38) while SNPs for celiac disease or rheumatoid arthritis were most enriched in DHSs from T-regulatory cells (67-72% SNPs, OR=3.4-4.9; p=0.002-0.04, H=0.23-0.38). In a new scan of eGFR (i.e. kidney function) among a total of >130,000 samples, genome-wide significant SNP associations were most enriched in DHSs from renal proximal tubule and cortical epithelial cells with ORs up to 3.7-7.1 (45-72% SNPs, p=0.02-0.001, H=0.30-0.64). Similarly, genome-wide significant SNP associations from a new combined metabochip+GWAS scan for blood pressure (BP) in up to 201,200 samples were most enriched in DHSs from microvascular endothelial cells with ORs up to 6.8-7.4 (76% SNPs, p=0.01-0.009, H=0.60-0.66). For both eGFR and BP, SNPs with larger association p-values showed less cell-type specific enrichment in DHSs compared with reference SNPs from the GWAS catalog. This trend may be interpreted as evidence for a lower proportion of true causal SNPs or associations with pleiotropic effects shared with other phenotypes among less significant SNPs. The proposed framework may be extended to other choices for reference SNPs or chromatin marks, e.g. H3K4me3 sites.

**1018T**

Down-regulation of the acetyl CoA metabolic network in visceral and subcutaneous adipose tissue of diabetic but not healthy obese individuals. H. Dharuri1, P.A.C. ’t Hoen1,2, J.B. van Klinken1, P. Henneman2, J.F.J. Laros3,4,5, M. Lips2, H. Pijl3, K. Willems van Dijk1,2, V. van Harmelen1. 1) Department of Human Genetics, Leiden University Medical Center, Leiden, RC, Netherlands; 2) Department of Clinical Genetics, DNA Diagnostics Laboratory, University of Amsterdam, Amsterdam, The Netherlands; 3) Department of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands; 4) Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands; 5) Netherlands Bioinformatics Center, Nijmegen, The Netherlands.

Obesity, a growing world-wide epidemic, is associated with decreased life expectancy due to associated metabolic and cardiovascular disorders. The expanded adipose tissue is thought to serve as the pathogenic link between obesity and type-2 diabetes. Furthermore, it has been shown that the accumulation of visceral adipose tissue (VAT) poses a greater metabolic risk than the subcutaneous adipose tissue (SAT). While a majority of obese individuals develop insulin resistance and type-2 diabetes, some remain metabolically healthy. We hypothesize that the biochemical mechanisms that underlie the function of adipose tissue can help explain the difference between healthy and diabetic obese subjects. To address this hypothesis, we utilized RNA-sequencing to analyze the transcriptome of samples extracted from VAT and SAT of obese female subjects undergoing bariatric surgery. A differential gene expression analysis between diabetic and non-diabetic individuals within the cohort pointed to important differences both in the VAT and SAT. Bioinformatic visualization techniques implicated acetyl coA metabolism. A workflow management software called Taverna was used to generate a gene set for acetyl coA reaction network using the KEGG database. Gene set analysis using the global test R package pointed to a statistically significant association with the phenotype. Multiple genes in the immediate vicinity of acetyl coA in the reaction network are down-regulated in diabetic obese subjects. To ascertain if the down-regulation of these genes is correlated to health status, we compared the pre and post-surgery (3 months after) expression levels of these genes in the SAT by qPCR. At this time the majority of diabetic obese women had a significantly improved metabolic health status. We observed a statistically significant up-regulation of acetyl-CoA acetyltransferase 1 (ACAT1), acetyl-CoA carboxylase alpha (ACACA), aldehyde dehydrogenase 6 family, member A1 (ALDH6A1), and methylenetetrahydrofolate dehydrogenase (MTHFD1) post-surgery in diabetic subjects, and the expression levels of these genes were down-regulated in non-diabetic subjects. These represent novel loci associated with type-2 diabetes and recovery in the VAT and SAT of obese subjects. Moreover, our data suggest that acetyl coA metabolic dysregulation plays a role in the pathophysiology of type-2 diabetes in obese subjects.
1019F  

**Primary Progressive versus Bout Onset Multiple Sclerosis: GWAS and pathway-network analysis.** G. Giacalone1, M. Sorosina1, V. Martellini1, M. Leone1, S. D’Alfonso2, G. Comi1, F. Martellini Boneschi1. 1) San Raffaele Scientific Institute, Milan, Italy; 2) University of Eastern Piedmont, Novara, Italy.

**Background:** Genetic contribution to the two different Multiple Sclerosis (MS) courses, Primary Progressive MS (PPMS) and Bout Onset MS (BOMS), has previously been hypothesized by epidemiological studies, but so far no single genetic marker association has been found. Aim: To identify single nucleotide polymorphisms (SNPs) associated to MS course by comparing PPMS and BOMS through a genome-wide association study (GWAS) and a pathway-network analysis approach. Methods: A total of 444 PPMS and 541 BOMS patients genotyped on two different platforms were merged on common autosomal SNPs (296,589). After applying standard quality controls, logistic regression analysis was carried out including age, age at onset, gender and the first principal component as covariates. A pipeline of multiple-pathway and network analysis was developed to move from SNPs through gene-wise p-values (p), to biological enriched terms/networks and potential disease-modifier genes. In the pipeline we used Gene Ontology and KEGG databases and the following bioinformatic tools: proxyGeneLD, MetaCoreTM, GOSTAT, Genecodis, Pathway-Express. Results: ProxyGeneLD program mapped 296,589 SNPs to 16,583 genes according to linkage-disequilibrium structure (HapMap CEU II 23hap18) and computed gene-wise adjusted p. Genes with adjusted p <0.05 (n=558) were prioritized for those directly interacting with each other (MetacoreTM protein interactions database) and the final list (n=218) submitted to multiple-pathway and network analysis. The first 3 hubs of the network composed of 218 genes were HIF1A, ETS1, NOTCH1, known to be involved in angiogenesis, immunity and myelination. Multiple-pathway analysis produced 32 significant terms, the majority of which related to immune functions (‘B cell mediated immunity’, ‘T cell receptor signaling’, ‘Leukocyte transendothelial migration’, ‘Cytokine production’), ‘Canonical Wnt/receptor signaling’, ‘L-glutamate transport’. Network analysis produced 11 significant networks, some of them enriched also in immune functions and Wnt signaling. Conclusions: In the comparison of PPMS vs BOMS patients, pathway and network analyses approach suggest a possible role of genes related to immune and cytokine signaling pathways and myelination processes, which should be further investigated in additional studies.

1020W  

**Genome-wide analyses of TCF7L2 DNA occupancy across multiple cell lines point to genetic networks underpinning complex traits.** M.E. Johnson1, J. Zhao2, J. Schug2, S. Deliard2, O. X. Yuan2, V. Guy1, J. Sainz2, K.H. Kaestner1, A.D. Wells1, S.F.A. Grant1, 1) Division of Human Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 2) Department of Genetics and Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), Spanish National Research Council, Santander, Spain; 4) Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA.

**Variation at the TCF7L2 locus is strongly implicated in the pathogenesis of type 2 diabetes (T2D) and cancer.** We previously mapped the genomic regions bound by TCF7L2 using ChiP-seq in the colorectal carcinoma cell line, HCT116, revealing a highly significant over-representation of GWAS loci associated with T2D and cardiovascular disease among the TCF7L2-bound targets. We carried out ChiP-seq in HepG2 hepatoma cells, where TCF7L2 is abundantly expressed, and leveraged these data against six additional cell lines for which genome-wide TCF7L2 binding is now available from ENCODE. Analyzing each cell line in turn, we found a strong evidence for enrichment of endocrine (including T2D), cardiovascular and cancer GWAS categories. Ingenuity pathway analysis detected enrichment of the ‘T2D mellitus signaling’ category in TCF7L2-bound genes in all cell lines except in HepG2 and HEK293 suggesting that neither the liver nor the kidney is a major target tissue for TCF7L2 influence in type 2 diabetes. In addition, we observed consistent under-representation of factors involved in beta-cell related functions within this ‘T2D mellitus signaling’ category but over-representation of members of the AKT pathway in other tissue types. Finally, we identified conserved pathways that are consistently bound by TCF7L2 across multiple cell lines were also highly significantly enriched for genes implicated by GWAS. As such, we strongly suggest that knowledge of TCF7L2 occupancy gives us insight into the genomic inferences in important pathways influencing the risk of several complex phenotypes.
1023W
Currairino syndrome - Searching for factors modifying expression of MNX1. I. Holm1, T. Monclair2, B. Stadheim1, K.L. Eiklid1. 1) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 2) Department of Pediatric Surgery, Oslo University Hospital, Oslo, Norway.

Currairino syndrome (CS), (OMIM#176450) is an autosomal dominantly inherited developmental disorder caused by incomplete rostral separation of the endo- and ectoderm. The condition is characterized by three main findings: a sacral bony abnormality, anal stenosis and a presacral mass. CS is associated with mutations in the MNX1 transcription factor gene (previously known as HLXB9) in familial cases. Penetration is reduced and expression is variable within families. In sporadic cases less than one third of the patients have detectable MNX1 mutations. Clinical signs and symptoms are similar in familial and sporadic patients. We have detected mutations in MNX1 in all affected members of four Norwegian families, but in none of the six sporadic cases. MNX1 is one of many transcription factors working together in regulation of embryonal development. The transcription factor is expressed early in foetal life and is especially important for pancreatic development and development in the sacral region. It is assumed to play a role in neonatal diabetes (Bonnefond et al., 2013) although patients with CS do not have diabetes. We wanted to examine if transcription factors known to interact with MNX1 could lead to defective expression of MNX1 and thereby causing clinical CS features in the non-familial cases. We searched for sequence variation in both translated and untranslated regions in the transcription factors already known to interfere with MNX1. We have so far looked at SOX17 (OMIM # 610928), SHH (OMIM # 600725), CDX1 (OMIM # 600746), ISL1 (OMIM # 600366) and PDX1 (OMIM # 600733). Preliminary results show no common SNPs shared by all six sporadic patients. We also searched for sequence variation in the highly conserved region upstream of the MNX1 to see if that could influence expression of MNX1 in non-familial CS patients. aCGH with SNP-microarray were used to look for other regions with factors or genes explaining the symptoms in non-MNX1 patients. The complete result will be presented in the poster.

1024T
Analysis of GTF2IRD1’s role in the craniofacial and neurological features of Williams-Beuren syndrome indicates an epigenetic control function. S.I. Palmer1, C.P. Canales1, P. Carmona-Mora1, F. Tomaseletti1, J. Widagdo1, A.C.Y. Wong2, G.W. Housley2, P. Kaur1, I. Smyth1, A-J. Han2, P.W. Gunning1, E.C. Hardeman1. 1) School of Medical Sciences, Cellular and Genetic Medicine Unit, University of New South Wales, Sydney, Australia; 2) School of Medical Sciences, Department of Physiology and Translational Neuroscience Facility, University of New South Wales, Sydney, Australia; 3) Epithelial Stem Cell Biology Laboratory, Peter MacCallum Cancer Centre, Melbourne, Australia; 4) Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia; 5) The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Australia.

Williams-Beuren syndrome (WBS) is a neurodevelopmental disorder resulting from a hemizygous microdeletion within chromosome 7q11.23 involving 28 genes. Genotype-phenotype correlations in patients with atypical deletions have mapped the typical craniofacial dysmorphologies and most aspects of the cognitive profile to a pair of genes that encode the evolutionarily-related transcriptional regulators GTF2IRD1 and GTF2I. The gene GTF2IRD1 is centrally located in the deletion and we have generated Gtf2ird1 knockout mouse lines that have phenotypes reflecting aspects of the human disease. Similar to WBS patients, knockout mice have large lips and this phenotype correlates with the pattern of Gtf2ir expression in facial tissue during mouse development. RNA-seq analysis of knockout lip tissue suggests defects of proliferation and differentiation that implicates GTF2IRD1 in the control of patterning and development in lip epithelium. Gtf2ird1 is also expressed in discrete brain regions and sensory organs that correlate with other abnormalities in the knock-out mice. WBS patients, including defects of motor coordination, hearing, exploratory drive and anxiety. Gene expression profiling in the corpus striatum of Gtf2ird1 knockout mice shows changes in the activity of a set of immediate-early genes, indicating possible roles in adult neuron maintenance and the control of experience-induced gene activity. To understand how this may work at the molecular level, our biochemical experiments have shown that GTF2IRD1 negatively auto-regulates its own allele though direct DNA binding and utilizes protein interaction surfaces to cooperate with other DNA binding proteins. These interactions are thought to be mediated by a specific set of chromatin modifying proteins that would explain its role in gene silencing. The GTF2IRD1 protein is subject to tight control of abundance and activity through SUMOylation and ubiquitination mediated proteasomal degradation. Together, these data support a role for GTF2IRD1 in the neurological and craniofacial features of WBS and suggest that it regulates gene expression via epigenetic mechanisms.

1025F
A Disease Module Captures Novel Candidate Genes and pathways for Asthma. A. Sharma1,5, J. Mence2,3, S. Ghiasian1,4, A-L. Barabasi1,2,3. 1) Center for Complex Networks Research and Department of Physics, Northeastern University, Boston, MA 02115, USA; 2) Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA; 3) Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA.

To understand various complex disease mechanisms, we not only require cataloging of disease genes by current high-throughput technologies, but detail of cellular components that are influenced by these genes and gene products. Here, we propose a novel concept of disease module that aims to identify the local neighborhood of cellular network for a particular disease phenotype. We illustrate the disease module detection algorithm by identifying the asthma disease module. Further, we validate the module for functional and pathophysiological relevance, using both computational and experimental approaches. Asthma module provides a deeper knowledge of the potential candidate genes and pathways that drive the pathological processes. On one hand, asthma module is enriched with disease-associated variations and on other hand; the impact of asthma specific drug is limited to the disease module. The emergence of GAB1 signaling pathway with both genetics and therapeutics (glucocorticoid) influence establishes the potential of disease module algorithm for dissecting the complex diseases.

1026W
The generation of sex-specific allergy networks: Validation that allergic disease has sexual dimorphic origins. J. Lasky-Su1,2, W. Qiu1,2, V. Carey3, S. Weiss1,2, B. Raby1,2, D. DeMeco1,2. 1) Channing Division of Network Medicine, Brigham and Women’s Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA.

Childhood allergy affects over 50% of children and has sexually dimorphic features that are evidenced in early childhood and into adolescence, including a higher prevalence of allergy and atopy in boys. A key characteristic of this sex difference is the consistently higher levels of total serum immunoglobulin E (IgE) observed in males. There is growing evidence that a portion of the underlying basis for the observed sexual dimorphism in allergic disorders is molecular, as heritability estimates for allergic phenotypes such as total serum IgE vary significantly for males and females (83% vs. 63%). Previously we examined the relationship between total serum IgE levels and genome-wide gene expression in 223 CAMP individuals and found striking differences by sex: 1) Sex stratified analyses revealed a significant correlation between IL17RB and IgE in males only (r^2 = 0.19, p value=8×10-8); 2) There were 873 compared to only 154 transcripts that mapped to gene IDs for males versus females; 3) Pathway enrichment analysis found nominal levels of enrichment in 13 pathways for males and only 5 pathways for females and none of these pathways overlapped. Given these notable differences between males and females, we constructed sex-specific allergy networks using gene expression data from the 223 CAMP individuals and applied Weighted Correlation Network Analysis (WGCNA). We first constructed sex-specific allergy networks using a general co-expression network based on the Pearson correlation coefficient by measuring the concordance of gene expression measures. These correlation measures are used to define an adjacency matrix, which indicates the degree that two expression measures are connected. We then identified important allergy and sex-specific modules within the networks using hierarchical clustering and topological dissimilarity. Once the sex-specific allergy networks were concordant of gene expression measures, these correlation measures are used to define an adjacency matrix for a particular disease. We then identified sex-specific biological pathways that contribute to allergy pathogenesis.
1027T Gene network analysis of candidate loci for human anorectal malformations. M. Garcia-Cardelio1,2,4, E.H.M. WONG3, C.H. NG3, V.C.H. LUI1,4, M.T. SO1, S.S. CHERNY2,3, P.C. SHAM1,2,4, P.K. TAM1,4, 1) Dept Surgery, Univ Hong Kong, Hong Kong, NA, Hong Kong; 2) Dept Psychiatry, Univ Hong Kong, Hong Kong, NA, Hong Kong; 3) Center for Genomic Sciences, Univ Hong Kong, Hong Kong, NA, Hong Kong; 4) Centre for Reproduction, Development, and Growth Univ Hong Kong, Hong Kong, NA, Hong Kong. Anorectal malformations (ARMs) are birth defects that require surgery and carry significant chronic morbidity. Our genome-wide copy number variation (CNV) study had provided a wealth of candidate loci. To find out whether these candidate loci are related to important developmental pathways, we have performed an extensive literature search coupled with currently available bioinformatics tools. This has allowed us to assign both genetic and non-genic CNVs to interrelated pathways known to govern the development of the anorectal region. We have linked 10 candidate genes to the WNT signaling pathway and 17 genes to the cytoskeletal network. Interestingly, candidate genes with similar functions are disrupted by the same type of CNV. The gene network we discovered provides evidence that rare mutations in different interrelated genes may lead to similar phenotypes, accounting for genetic heterogeneity in ARMs. Classification of patients according to the affected pathway and lesion type should eventually improve the diagnosis and the identification of common genes/molecules as therapeutics targets.

1028F Comparison of gene expression induced by HIV-1 GAG peptides specific to HLA-A*01:01 and B*07:02 in PBMCs by mRNA-seq analysis. L.R. Liu1,2, P. LaCap1, R. Capina1, B. Liang1,2, B. Fristensky1,3, B. Ball1,2, F. Plummer1,2, M. Luo1,2, 1) HIV and Human Genetics Division, National Microbiology Laboratory. 1015 Arlington Street, Winnipeg, Manitoba, Canada R3E 3R2; 2) Department of Medical Microbiology, University of Manitoba. 745 Bannatyne Avenue, Winnipeg, Manitoba. R3E 0J8; 3) Department of Plant Sciences, University of Manitoba. 222 Agriculture Building University of Manitoba, Winnipeg, Manitoba. R3T 2N2. Introduction: A subset of sex workers (CSEW) enrolled in the Pumwani cohort in Nairobi, Kenya remain HIV negative despite repeated exposure through high risk sex work. Studies on genetic factors enriched within these highly HIV-1 exposed seronegatives (HESN) suggest this natural resistance to HIV-1 is multi-factorial, and associated with specific alleles of Human Leukocyte Antigens (HLA). HLA-A*01:01 was associated with reduced risk of HIV-1 infection, whereas B*07:02 was associated with increased risk. HLA initiate cell-mediated immunity (CMI) by presenting antigens to T-cells. Systematic comparison of A*01:01 and B*07:02 HIV-1 Gag epitopes showed that A*01:01 recognized fewer epitopes than B*07:02, and recognition of more Gag epitopes is associated with susceptibility to HIV-1 infection. However, it is unclear whether the A*01:01 or B*07:02 and GAG peptide complex could differentially induce downstream gene expression, and lead to differential T cell function. We conducted mRNA-seq analysis and compared gene expression induced by A*01:01 and B*07:02 specific peptide using PBMCs of an individual express both A*01:01 and B*07:02. Approach: PBMCs from a single blood draw was split into equal proportions, and stimulated with peptide NSSKVSQNY (A*01:01 specific) or SPRTLNAWV (B*07:02 specific). 8-hours post-stimulation, the PBMCs were harvested for RNA-seq analysis. RSEM (RNAseq by Expectation Maximization) workflow was used for read alignment, and transcript quantification. DESeq was used for differential expression analysis. Results: After correction for false-discovery, one gene was significantly differentially expressed, DEOS71957 (FDR-P-value < 3.8 x 10^-15), and 190 gene-isofoms (FDR-P-value ≤ 0.05). Implications: This study aims to inform vaccine design by evaluating host gene expression induced by HLA-specific epitopes. Significant differences were identified however the results need to be validated by qRT-PCR and in other A*01:01 and B*07:02 co-expressed individuals.

1029W Protein altering variants found in ciliary and polarity genes in biliary atresia patients. E.A. Tsar1, M. Grochowski2, L.D. Leonard2, R.P. Matthews3,4, K.M. Loome3,4, B.A. Habib5, N.B. Sperner6,5, M. Devoto7,8,9 1) Genomics and Computational Biology Graduate Group, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA; 4) Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA; 5) Hepatology, Infectious Diseases Clinical Research Department, Merix Research Laboratories, North Wales, PA; 6) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia PA; 7) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 8) Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 9) Department of Molecular Medicine, University La Sapienza, Rome, Italy. Biliary atresia (BA) is a bile duct disorder that presents within the first few months of life and causes necroinflammatory obliteration of the extrahepatic biliary tree. Children with BA have severe liver disease, and BA is the most frequent indication for pediatric liver transplantation in the United States. BA is thought to result from a combination of genetic and environmental risk factors, but no specific gene responsible for BA has been identified yet. Absence of immotile cilia on the surface of endothelial cells is apparent upon inspection of intrahepatic as well as remnant extrahepatic bile ducts, suggesting a loss of cell polarity. Manipulation of polarity genes leads to biliary defects in model organisms, and some BA patients demonstrate other anomalies consistent with polarity defects. We compiled a list of genes that participate in the establishment or maintenance of cell polarity (n=280) as well as genes with a role in the composition or function of cilia (n=291). We hypothesize that rare or novel damaging mutations in these genes may contribute to the development of BA. We performed exome sequencing with the Agilent SureSelect All Exon V4+UTR capture kit on 30 Caucasian, isolated BA patients. Variant filtration was performed to analyze only variants with frequency ≤5% in the 1000 Genomes Project Phase I and the NHLBI Exome Sequencing Project (ESP). Notably, one male patient had a nonsense variant in the X-linked polarity gene, ATP6AP1. This variant, rs201620814, was not observed in 1000 Genomes, but heterozygous variants in 4/1949 females and a hemizygous variant in 1/1283 males in the ESP cohort (<0.1%) was observed. Additionally, mutations in vacuolar H+-ATPase subunits have been shown to cause biliary defects in zebrafish. Another patient had two missense variants in trans in DNA1, which encodes for part of the outer dynein arm of cilia. We have identified several other changes in polarity and cilia genes that we suspect are contributing to BA in these patients. Our results support the hypothesis that polarity and cilia genes may be important in BA etiology but also suggest a high level of genetic heterogeneity in this disorder.
1030T
Vascular tone pathway in relation to primary open angle: results from the NEIGHBORHOOD consortium. L. Pasquale1, J. S. Loomis1, J. Kang2, J. Bailey2, J. Haines3, B. Yaspan1, M. Hauser1, J. Wiggs1. NEIGHBORHOOD consortium. 1) Glaucoma Service, Massachusetts Eye & Ear Infirmary, Boston, MA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 3) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 4) Duke University School of Medicine, Durham, NC.

Background: Reduced ocular perfusion pressure and impaired autoregulation of ocular blood flow have been implicated in the pathogenesis of primary open angle glaucoma (POAG). Genome wide association studies (GWAS) of POAG have identified CAV1/CAV2, which codes for caveolins that are involved in vascular regulation. Biologic pathway analyses can enhance the power to discover biologically meaningful genetic markers associated with complex traits. We investigated whether a set of genetic markers of factors involved in setting vascular tone are associated with POAG. Methods: We used Illumina 660W-Quad array platform genotype data and pooled p-values from 3,108 POAG cases and 3,430 controls from the combined the National Eye Institute Glaucoma Human Genetics Collaboration consortium and the Glaucoma Genes and Environment study. We compiled a set of single nucleotide polymorphisms (SNPs) in 186 genes related to vascular tone that were determined with KEGG pathways or previously identified in GWAS for blood pressure. We analyzed the association between these SNPs and POAG using both a pathway- and gene-based approach with the Pathway Analysis by Randomization Incorporating Structure analysis software package, which performs a permutation algorithm to assess the statistical significance of gene and pathways relative to randomly generated genes and pathways of comparable size, structure and linkage patterns. Results: The vascular tone pathway was not associated with POAG overall or the POAG subtypes defined by type of visual field loss (paracentral [n=224 cases] or peripheral loss [n=993 cases]) (permuted p=0.20). In addition, results were similar null in men and women (permuted p=0.81). In gene-based analyses, two genes had nominally significant permuted gene p-values <0.05 in almost all analyses by subtype and gender: CAV1/CAV2 (permuted gene p=0.01 for all outcomes), PRKAA1 (permuted gene p=0.05 for all outcomes); in addition, TNF alpha encoding (permuted gene p=0.01 for all outcomes). Conclusions: Although the overall vascular tone SNP set did not show significant associations with POAG, the gene-based analyses confirm the previously published association of the RXFP3 gene and suggest a possible role of caveolins and related factors in POAG pathogenesis.

1030F
Polygenetic Risk Model Suggests a Protective Effect of Lipid Genes on Plasma Glucose and HbA1c levels. J. Fu1, N. Liu2, M. van der Siddeel1, C. Wijmenga1, H. Snieder2, M. Hofker3. 1) Genetics Dept, UMCG, University of Groningen, Groningen, Netherlands; 2) Molecular Genetics, UMCG, University of Groningen, Groningen, Netherlands; 3) Epidemiology Dept, UMCG, University of Groningen, Groningen, Netherlands.

Background and hypothesis: Dyslipidemia is strongly correlated with raised glucose levels and type 2 diabetes (T2D), although the causal nature of the correlation and the role of dyslipidemia genes on T2D are still unclear. The advances made by genome-wide studies are allowing the use of a polygenetic risk score to assess the impact of lipid genes on T2D or related traits. Methods: We studied 95 common genetic variants that were robustly associated with triglyceride (TG), total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C) or high-density lipoprotein cholesterol (HDL-C). We included 13,107 subjects from the LifeLines cohort and calculated the polygenetic scores of blood lipids per individual. We computed the Spearman correlation between each genetic score and fasting plasma glucose and HbA1c levels. Results: Despite the positive correlation between dyslipidemia and HbA1c, surprisingly the individuals carrying a higher triglyceride genetic risk seemed to have lower HbA1c levels (r=0.025, p=0.009). Conditional on blood lipid levels, this negative correlation became not only stronger for HbA1c (r=0.043, p=2x10-6) but also significant for fasting glucose level (r=-0.051, p=8.6x10-8). We also consistently observed that the higher genetic scores of TC and LDL-C were correlated with lower HbA1c levels (p=7.2x10-4 and p=0.003, respectively). Conclusions: In the LifeLines cohort, our polygenetic risk model suggests that genes associated with dyslipidemia may have a weak protective effects on plasma glucose and HbA1c levels. This protective effect is pleotropic and independent of all types of lipid.

1032W
Molecular intermediate phenotype mapping of IL-6 and TNFα levels reveals genes critical for chronic systemic inflammation. S. J. Schrodi1, D. David2, J.K. Meece3, D.A. Vasco1, J.J. Mazza4, M.H. Brilliant1, J.A. Smith5. 1) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 2) Core Laboratory, Marshfield Clinic Research Foundation, Marshfield, WI; 3) Clinical Research Center, Marshfield Clinic Research Foundation, Marshfield, WI; 4) Department of Pediatrics, University of Wisconsin-Madison, Madison, WI.

Background Numerous common diseases exhibit an underlying component of chronic systemic inflammation. Discovery of genetic variants that produce susceptibility to these diseases has many pitfalls, not the least of which lies in the clinical definition of disease which may or may not reflect pathological mechanisms at the molecular level. Hence, redefining phenotypes based on molecular characteristics may elicit more robust and lucid genetic association signals. This study uses a composite metric of the circulating levels of two important and correlated inflammatory cytokines, TNFα and IL-6, in the general population as a molecular intermediate phenotype. Both cytokines are important in the immunobiology of proinflammatory processes and have been directly targeted by biological therapeutics to remedi- ate autoimmune conditions. To identify novel genes involved in regulation of TNFα/IL-6-mediated chronic inflammation, we performed a quantitative trait GWAS on TNFα/IL-6 levels in a genetically homogeneous population in Central WI, largely derived from Bavarian immigrants in the late 1800s.

Materials/Methods Using the Meso-scale Discovery platform, we measured levels of TNFα and IL-6 in plasma in 135 individuals not taking immune-modulating therapeutics or having evidence of acute infection. A small number of genetic background outliers were removed from the sample set. The Illumina 660W-Quad array was applied to all samples and a linear correlation test was applied for the genotypes at each of >500,000 SNPs to transform TNFα/IL-6 concentration data. The VEGAS program was applied to individual SNP data to obtain gene-based association P-values.

Results/Conclusions From roughly 18,000 genes evaluated, two statistically compelling regions emerged from the analysis: 1) the RXFP3-SLC45A2 region on the short arm of Chr5 (P<1E-06), and 2) the GPR31-CCR6 region on the long arm of Chr6 (P<1E-06). RXFP3 encodes for a relaxin/insulin-like receptor, and may wield anti-fibrotic, and anti-inflammatory actions. SLC45A2 encodes for a melanocyte differentiation antigen and is intimately involved in melalin synthesis (OMIM). Interestingly, Sellick et al (2005) mapped the GPR31-CCR6 region as conferring strong predisposition to small vessel lymphocytic vasculitis within an extended family. Additionally, CCR6 encodes for a chemokine receptor expressed on lymphocytes. These results may provide unique insight into the molecular mechanisms behind proinflammatory processes.
1033T
Uncovering the genetic architecture of complex traits using network approaches. *J. Choi*, C. Cotatsap9,10,11,12,13,14. 1) Department of Neurology, Yale School of Medicine, New Haven, CT; 2) Department of Genetics, Yale School of Medicine, New Haven, CT; 3) Medical and Population Genetics, the Broad Institute of Harvard and MIT, Cambridge, MA.

Over the last decade, genome wide association studies (GWAS) have uncovered thousands of genetic variants influencing every aspect of all organ systems. However, it has proven difficult to infer which genes these variants are perturbing, so the challenge of uncovering underlying biological mechanisms remains elusive. We do not yet understand the genetic architecture of traits, what is the true performance of networks representing pathways [Rossin et al PLoS Genetics 2011]. Here, we project GWAS scores onto protein-protein interaction networks and use robust network clustering algorithms to look for regions of the overall interaction network enriched for disease association signal, indicating the presence of an interacting set of genes modulating susceptibility. Within this framework, we investigate several genetic architecture models by simulating scenarios where either a single or multiple groups of interacting genes modulate risk versus no aggregation of signal in gene groups. We are able to recover both true positive and false negative associations, increasing the heritability explained by GWAS. Surprisingly, we also find that in models of multiple interacting pathways as a single large, connected gene network. This appears to be a property of gene interaction networks and suggests that recent reports of large sets of interacting genes underlying disease susceptibility are in fact capturing multiple biological pathways. In two disease GWAS meta-analyses (Crohn’s disease and multiple sclerosis) we find gene networks of size ~100 enriched for genetic risk, consistent with an architecture of cumulative genetic burden on molecular pathways. We also find that subsets of these components are expressed in different tissues, indicating that both disease and tissue are interconnected networks. A logistic regression path model identifies influential topological features and enables prediction of tissue-disease relationships represent known pathogenesis; and disease-specific AUC in the range of 0.70-0.79. Using our predictions as prior probabilities, we identified candidate genes that have not been associated to date. Relying on independent GWAS for discovery and validation, we highlight REL, JAK2, and TNFAIP3 as likely candidates in MS. The probability of the observed validation rate occurring by random prior estimates is 0.002, underscoring our ability to uncover novel associations. Integrating information across multiple domains through the use of heterogeneous networks significantly improved performance. Topological analysis of heterogeneous networks provides a potentially powerful new platform for discovery in human genetics.

1035W
Novel approach using gene set enrichment based on human genome-wide association study data implicates FXR/RXR activation as a common pathway affecting blood lipids levels and nonalcoholic fatty liver disease, Y.M. Puentes1,2, C.C. Powell1,2, L.M. Yerges-Armstrong1,2, M.F. Feitosa4, L.F. Bielak5, A.V. Smith5, T.B. Harris5, J. Liu6, S.K. Musani7, I.B. Boracci8, P.A. Peyser9, E.K. Spielotes1,2, GOLD Consortium: 1) Department of Bioinformatics and Computational Biology, University of Michigan, Ann Arbor, MI; 2) Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan 48109, USA; 3) Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA; 4) Division of Statistical Genomics, Department of Genetics, Washington University, St. Louis, Missouri 63108, USA; 5) Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor Michigan, 48109, USA; 6) Icelandic Heart Association, Kopavogur IS-201, Iceland; 7) Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA; 8) University of Mississippi Medical Center, Jackson, Mississippi, 39213, USA.

Nonalcoholic fatty liver disease (NAFLD) is rapidly increasing in prevalence and will become the number one cause of liver disease worldwide by 2020. NAFLD is associated with high triglycerides (TG), high low-density lipoprotein cholesterol (LDL-C) levels, and low high-density lipoprotein cholesterol (HDL-C) levels. Both NAFLD and blood lipid levels are genetically influenced and may share a common genetic etiology. We aimed to identify genes and pathways enriched for genetic associations with both blood lipids and NAFLD using human genome wide association study (GWAS) data. We examined novel approaches using geneset enrichment analysis (GSEA) as implemented in MAGENTA to identify pathways enriched for correlations between lipids, known NAFLD GWAS associations and these enriched pathways were then tested for enrichment in the NAFLD GWAS (N=7,126). We then used gene set enrichment analysis (GSEA) as implemented in MAGENTA to identify pathways enriched for correlations between lipids, known NAFLD GWAS associations and these enriched pathways were then tested for enrichment in the NAFLD GWAS (N=7,126). These enriched pathways were then tested for replication in an independent NAFLD GWAS (N=3,124, Palmer 2013). We found that TG (P=0.004) and LDL-C (P=0.036) were significantly enriched in the NAFLD GWAS. We identified 58 pathways that were enriched in lipid GWAS data. Three of these were also enriched in the NAFLD GWAS (N=7,126) and one, FXR/RXR activation, also showed significant enrichment in another NAFLD GWAS (N=3,124). None of these enriched pathways were enriched for associations in control publically available GWAS analyses of diastolic and systolic blood pressure (N=275,000), body mass index (N=249,796), and waist to hip ratio (N=77,167 suggesting that the enrichment was specific to NAFLD. Genes associated with NAFLD (P<0.05) in FXR/RXR activation fell into three functional categories: (1) VLDL Assembly: MTP, APOA1, ACBP3 (2) Nuclear Related Processes: PPAR-alpha, HNF1a, NR0B2/SHP and (3) Hepatic Transport: MRFP2, AE2, ABCG8, ABCG5, OAT2. Using a novel approach, we found that human genetic variation in or near genes involved in FXR/RXR activation affects both blood lipids and NAFLD in humans. These results suggest that genes that play a role in lipoprotein assembly, nuclear receptor biology, and hepatic transport when altered may affect NAFLD and thus could provide possible therapeutic targets for NAFLD prevention or treatment.
1036T


The identification of a functional association among a set of genes whose variants are implicated in the same disease can suggest cellular processes relevant to the disease mechanism. As the function of many genes is not known and their classification to pathways is scant, functional links between genes are often inferred from large-scale `omics` data.

We have derived functional associations between human genes from diverse data types and examined their disease relevance. Annotation data were processed in the form of semantic similarity, which is a measure of relatedness between two genes as assessed by the similarity of their annotations. We combined the most informative input data sources to form a functional-linkage network, weighting the different data types in proportion to their relative accuracy. Applying our framework, we have examined the network properties of sets of genes with de novo substitutions implicated in the same disease by exome sequencing studies.

We found that mouse knockout phenotypes were the most informative predictors of functional links between human genes. Accordingly, genes mutated in the same disease clustered most significantly in a gene network built on mouse phenotypes and in the integrated functional-linkage network, while genes mutated in controls did not form clusters. Combining three sets of genes implicated in autism by recent exome sequencing studies, we have observed distinct functional sub-clusters in the network, suggesting relevant biological pathways.

Functional associations between genes are often inferred from omics data, but the disease relevance of such data types, including protein-protein interactions and gene co-expression networks, has been unclear. The application of semantic similarity has enabled us to compare diverse data types and establish the value of mouse phenotypes. We present a framework for the identification of biological pathways disrupted in complex disorders and demonstrate the ability of this approach to functionally dissect the molecular variants underlying autism.

1037F

Gene-set enrichment analysis of gene expression associations with acute asthma control hints at candidate drug pathways. D.C. Croesuch-Chonka1, W. Qui1, V.J. Carey4, S.T. Weiss2, B.A. Raby1,3, Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE) Consortium. 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Partners Center for Personalized Genetic Medicine, Partners Health Care, Boston, MA; 3) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Maintenance of effective asthma control is critical for preventing exacerbations, or sudden declines in lung function. Loss of control results in considerable economic costs due to chronic treatment, acute hospitalization, and/or absence from work or school. By better understanding the biological underpinnings of acute asthma control (AAC), we can develop improved clinical and pharmaceutical approaches for maintaining effective disease stability.

To understand the relationships of gene expression with AAC, we performed differential expression analyses in a subset of 583 subjects from the Asthma BioRepository for Integrative Genomic Exploration (BRIDGE) cohort (62% male, mean age = 8.9 years) who had genome-wide gene expression data derived from whole blood RNA extracts and measured with the HumanHT-12 v4 Expression BeadChip. AAC was represented by a modified version of the clinically validated Asthma Control Test (mACT) score. After non-specific filtering, 10,701 probes (one per gene) were tested for linear associations with the normally distributed mACT, adjusting for batch effects using the first ten principal components of gene expression. A total of 548 probes were nominally associated with mACT (P < 0.05), but none were associated at a false-discovery rate (FDR) < 0.05.

To gain further biological insights, we performed a gene-set enrichment analysis (GSEA) of the ranked associations using the GSEA2 software and the `C2` collection of 3,112 curated genes from the Molecular Signatures Database (MSigDB). We observed significant enrichment (FDR < 0.25) of 288 MSigDB gene sets among genes whose increased expression was associated with below-median mACT and two gene sets associated with above-median mACT. At an FDR < 0.05, the counts were 41 and zero, respectively (mean gene set size = 62). Enriched gene sets included ones related to expression in immune cells (e.g. plasma cells, eosinophils, B lymphocytes) and to changes in expression in response to non-chemotherapeutic drugs (e.g., glucocorticoids, prostaglandin E2, methoxysteradiol).

These preliminary results suggest the mACT phenotype is a biologically meaningful representation of AAC and hint at potential candidate pathways for drug treatment development. This work was supported by NHLBI grants RC2 HL101543, R01 HL086601-6.

1038W

Gene set signature of leprosy reversal reaction type 1. M. Orlova1, A. Cobat1,2, N. Thu Huong2, N.N. Ba1, N.V. Thu2, J. Spence1, Y. Nédélec2, L. Barreiro3, V.H. Thao2, L. Abecasis1,6,7,8, A. Alcais4,6,9, E. Sobry1,2,1. 1) The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 2) Departments of Human Genetics and Medicine, McGill University, Montreal, Quebec, Canada; 3) Hospital for Dermato-Venerology, Ho Chi Minh City, Vietnam; 4) Department of Microbiology, Immunology and Pathology, College of Veterinary Medicine & Biomedical Sciences, Colorado State University, Fort Collins, CO; 80523-1619, USA; 5) Department of Pediatrics, Sainte-Jeanne d'Arc Hospital Research Centre, University of Montreal, Montreal, Quebec, H5T 1C5, Canada; 6) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale, U980, Paris, France; 7) University Paris Descartes, Imagine Institute, Paris, 75993, France; 8) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10165, USA; 9) URC-CIC, Hospital Tarnier, Paris, 75006, France.

BACKGROUND. While effective drug treatment of leprosy is widely available, some leprosy patients suffer from major nerve damage due to leprosy type 1 reversal reactions (T1R). T1R are sudden episodes of exacerbated local delayed-type hypersensitivity to Mycobacterium leprae in skin and/or nerves. The incidence of T1R varies widely from 6% to 67% of leprosy patients in different leprosy endemic settings. T1R is effectively treated by corticosteroids, yet neurological impairment persists in about one third to half of all cases. Thus, early identification of patients at risk of T1R, and consequently at risk of neurological injury, is a major challenge in leprosy care.

METHODS. We used whole blood cultures stimulated with high doses of M. leprae sonicate to assay transcriptional responses of leprosy patients. To avoid the problem of highly dysregulated inflammatory responses during acute episodes of T1R we enrolled cured leprosy patients who had undergone T1R more than five years before and compared their transcriptional responses of M. leprae antigens at the time of enrolment to controls who developed T1R and those who did not. RESULTS. We defined and replicated prospectively a T1R signature set of 44 differentially regulated genes. Functional enrichment analysis of the T1R signature demonstrated that the majority of the T1R set genes: a) pro-inflammatory regulators; b) arachidonic acid metabolism mediators and c) regulators of anti-inflammation. The validity of the T1R gene set signature in the prospective arm directly demonstrated that T1R patients have an innate defect in the regulation of inflammatory responses to M. leprae. The T1R gene set appeared as molecular characteristic of T1R susceptibility and was consistent with a breakdown of communication between pro- and anti-inflammatory responses in T1R patients. The identification of the T1R gene set represents a critical first step towards a genetic profile of leprosy patients who are at increased risk of T1R and concomitant nerve damage.
pulmonary function measures in our pediatric Caucasian cohort and the GWAS. We found that the glycoprotein gene set is associated with all four pulmonary function traits tested. In our study, we observed significant responses to LPS for 6,021 genes expressed with p<0.05 in at least 20% of samples. At a false discovery rate (FDR) of 10%, we observed significant responses to LPS for 5,441 and 4,034 genes in non-atopic and atopic children, respectively. 841 of those genes showed significant response in atopic children only, and were enriched for genes in immunological disease pathways (p=0.0015), inflammatory response pathways (p=0.039), and pathways in pat- tern recognition receptors of bacteria and viruses (p=0.025). Overall, periph- eral blood leukocyte response to LPS identified significant differences in the response in individuals with asthma or atopy that may underlie some of the clinical features of these conditions. This work was supported by NIH grant HL085197.

### 1040F Gene Network Analysis with GWAS Data Identifies Novel Lung Function Gene Set

J. Li, B.A. Ong, J.M. McDonough, Z. Wei, C. Kim, R. Chiavacci, F. Mentch, J.B. Caboott, J. Spergel, P.M.A. Steinman, H. Hakonarson, 1 Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, PA; 2 Division of Pulmonary Medicine and Cystic Fibrosis Center, The Children’s Hospital of Philadelphia, Philadelphia, PA; 3 Division of Pediatric Pulmonology, Tripler Army Medical Center, 1 Jarrett White Rd, Honolulu, HI; 4 Department of Pediatrics, New Jersey Institute of Technology, Newark, NJ; 5 Division of Pediatric Pulmonology, Madigan Army Medical Center, Tacoma, WA; 6 Center for Pediatric Eosinophilic Disorders, The Children’s Hospital of Philadelphia, Philadelphia, PA; 7 Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA; 8 Division of Human Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA.

Pulmonary function reflects various pulmonary conditions and predicts morbidity and mortality in adults. While several studies have already been published on pulmonary function traits, these have mainly been carried out in adults of Caucasian ancestry leaving the genetic determinants of pulmonary function traits in the pediatric population and in African Americans, in particular, relatively unexplored. To identify new genetic determinants of pediatric lung function, we conducted a genome-wide association study (GWAS) of four pulmonary function traits, including FVC, FEV1/ FVC and FEF25-75. In 1556 children. Though no SNP reached genome-wide significance, we identified SNPs with notable trend towards association with the pulmonary function measures in both Caucasian and African American cohorts. We also replicated the previously reported INTS12 locus association with FEV1 (P<1.4x10^-7). We subsequently carried out gene network analysis for each trait using SNPs with P-value of < 1.0 x 10^-5 in GWAS. The gene expression was correlated with all four pulmonary function measures in our pediatric Caucasian cohort and the association was validated in our pediatric African American cohort and two reports. The meta-analysis p-values range from 6.29x10^-7 to 2.80 x 10^-8 for the four pulmonary function traits for our study, we provide the first comprehensive analysis of pulmonary function traits in pediatric population and African Americans and identified novel pulmonary function specific gene set in that population.
Psoriasis susceptibility genes in patient with severe compared with mild phenotype. P. Nikamo, M. Ståhle. Department of Medicine, Dermatology Unit, Karolinska Institutet, Stockholm, Sweden.

Psoriasis is usually graded as mild (affecting less than 3% of the body), moderate (affecting 3-10% of the body) or severe (affecting more than 10% of the body). According to National Psoriasis Foundation the distribution of psoriasis severity are 67% mild, 25% moderate and 8% severe (http://www.psoriasis.org/). The degree of severity is generally based on the following factors: the proportion of body surface area affected; disease activity (degree of plaque redness, thickness and scaling); response to previous therapies; and the impact of the disease on the person. Several psoriasis GWAS have been published, but phenotypic stratifications are lacking. Our large patient material with detailed clinical and genotype information allows for different stratification procedures. Healthy controls (n=1750) are compared with patients with mild psoriasis (only topical therapy n=712) and severe psoriasis (systemic therapy n=543). The analysis involves genes known to associate with psoriasis. Major differences are detected, such as only patients with severe psoriasis associate with the IL23R (p=0.0007, OR=1.31, CI 1.12-1.53) and IL23A (p=0.0001, OR=2.04, CI 1.40-2.93) - genes in a key psoriasis pathway (the effective biologic ustekinumab is directed against IL23A/IL12). In contrast, HLA-C and its putative functional partner ERAP1 (peptidase involved in trimming MHC Class I peptides) associates mostly with mild phenotype. These novel and other dramatic data and additional genotypes are now being explored to dissect the differences between these phenotypes. Altogether, these findings indicate that genes participating in IL23 signaling play a significant role in the severity of psoriasis. Conversely, the characterization of non-MHC disease loci has been problematic, owing to the small effects of the underlying genetic variants. This can be overcome by careful phenotyping and characteristics of the material.

Validating trans-eQTLs using evidence of cis-mediation: a genome-wide analysis among 1,800 South Asians. B. Pierce1, L. Tong1, R. Rahaman1, L. Chen1, M. Kibria1, M. Argos1, J. Farzana1, R. Shantanu1, R. Paul-Brutus1, R. Zaman1, M. Rahman1, J. Baron1, H. Ahsan1, 1) Health Studies, Univ Chicago. Chicago, IL; 2) UChicago Research Bangladesh, Dhaka, Bangladesh; 3) University of North Carolina, Lineberger Comprehensive Cancer Center, Chapel Hill, NC.

Genome-wide studies of expression quantitative trait loci (eQTLs) indicate that the majority genes in the human genome show evidence of regulation by a cis-eQTL (i.e., cis-eSNP). However, less is known regarding the effects of genetic variants on expression of distant genes (trans-eQTLs), which are difficult to detect in genome-wide analyses for two primary reasons: (1) trans effects, on average, are weaker than cis effects and (2) a huge number of tests must be conducted to comprehensively assess the evidence for trans eQTLs, resulting in the use of stringent significance thresholds. Several large eQTL studies (>1,000 participants) have attempted to identify trans signals, typically using leukocytes as source of RNA. These studies have reported hundreds of ‘significant’ trans-eQTL signals, but only a fraction of these signals have been identified across multiple studies. While there are many explanations for this lack of consistency for trans findings, this observation highlights the difficulty of identifying true trans signals with confidence. In this work, we use data on genome-wide SNP and mononuclear cell array-based expression data from ~1,800 Bangladeshis individuals to (1) characterize the cis- and trans-eQTL patterns among South Asians and (2) demonstrate trans signals can be validated by assessing evidence for cis-mediation. We first used standard eQTL methods to identify cis and trans signals using a false-discovery rate (FDR) of 0.05. Among the 448 independent putative trans-eSNPs we identified, we could classify 52 of these as cis-eSNPs (defined as 12 >0.8 with the lead eSNP for a known cis-eQTL signal). For each of these 52 SNPs showing both cis and trans associations, we estimated the trans association both with and without adjustment for the cis-transcript, a potential mediator of the trans effect. For approximately half of these signals, the magnitude of the trans-eQTL association was dramatically reduced (fold change >0.5) and in four cases, implying that the trans effect is mediated by a measured cis-transcript. For weaker trans signals that did not surpass the FDR threshold, we show that cis-mediation analysis can identify a subset of these signals that are likely to be true effects. In conclusion, assessing evidence for cis-mediation is a promising and straightforward approach for validating trans signals and enhancing trans-eQTL discovery. This approach may be especially useful when ideal replication data is not available.

Transcriptome analysis of CD4+ lymphocytes in asthmatics with or without depression. J. He, T. Wang, Z. Liang, X. Xiong, Y. Yang, Y. Ji. Department of Respiratory, Sichuan University, Chengdu, Sichuan, China.

Rationale: Cumulative studies have shown that asthma is positively associated with depression. The underlying mechanisms are poorly understood. Assuming common pathophysiological mechanisms especially immune mechanisms exist, the aim of this study was to determine the common pathways between asthma and depression using the global gene expression pattern of CD4+ lymphocytes from asthmatics with or without depression. Methods: Four groups of subjects (Non-depressive asthmatic, NDA; Depressive asthmatic, DA; Depression, DE; and Healthy control, HC) consisted of 6 participants in each group were studied. Peripheral CD4+ lymphocytes were isolated and the global transcriptome profile was performed using the Agilent SurePrint G3 Human GE 8x60K Microarray. Experimental data analyses were performed using the software R and Bioconductor. Genes with expression level of at least ± 2.0-fold change and a false discovery rate ≤ 0.05 were defined as statistically significant changes. Functional analysis and pathway enrichment analyses of differentially expressed genes were performed using the Ingenuity Pathway Analysis (IPA). Results: A total of 1145 differentially expressed genes were identified in any of the AS vs. HC, DA vs. HC, or DE vs. HC comparisons after corrected for multiple comparisons. Among these genes, 148 genes were demonstrated as differentially expressed genes only in DA vs. HC. Gene pathway analysis using these 148 genes identified six significant biological processes: Acute Phase Response Signaling, Nicotine Degradation III, Melatonin Degradation I, Superpathway of Melatonin Degradation, Serotonin Degradation and Nicotine Degradation II. Conclusion: The use of microarray analysis to identify the small but differentially expressed CD4 T cells in non-depressive asthmatic; depressive asthmatic and depression groups may provide a foundation for further studies on the pathogenesis of co-exist of asthma and depression. The results of this study shed new light on the common molecular mechanisms of asthma and depression and theoretical clues for further pathophysiological and clinical therapeutic studies.


The endoplasmic reticulum (ER) is responsible for synthesis and maturation of many proteins essential for cellular function. ER dysfunction can have devastating consequences when misfolded proteins accumulate in the ER, resulting in "ER stress" associated with pathogenesis of diseases from diabetes to neurodegeneration. To understand the contribution of ER stress resistance to disease, it is important to determine the extent of natural variation in ER stress resistance. Previous studies in human cell lines and Drosophila demonstrate that there is extensive variation in ER stress response, but comparable data in mouse is conspicuously absent. To characterize the variation in ER stress transcriptional response in mouse, we studied the eight genetically diverse inbred laboratory mouse strains that had been employed in the Collaborative Cross. Mouse embryonic fibroblasts (MEFs) from each strain were exposed to tunicamycin, an ER stress inducing drug. RNA-seq was performed to identify transcriptional changes in response to ER stress by contrasts with RNA-seq runs of cells under drug-free control conditions. Many canonical ER stress genes showed inter-line differences in response, indicating that the core ER stress pathways may be more plastic than previously thought. To partition this regulatory variation into cis and trans effects, we measured ER-stress-induced gene expression by RNA-seq in MEFs from F1 hybrids of the original eight inbred lines, and quantified allele-specific expression across nearly every expressed gene. The results reveal a surprising degree of trans regulatory variation, consistent with the variation being close to the head of the regulatory cascade. Whether the most important ER stress genes modulating variability in disease risk play these key regulatory roles remains to be determined. But differences in cis vs. trans regulatory architecture of genes in a network such as ER stress may have important implications for effect sizes of allelic variants, the nature of mutation-selection balances, and the prospects for therapies that target single gene products.
Global Metabolite profiling and the Risk of Osteoporotic Fractures: A Systems Biology Approach by Integrating Genomics and Tissue-Specific Gene Expression Profiling, Proteomics and Metabolomics in Postmenopausal Caucasian Women. YH. Hsu1,2, S. Reppe3, CL. Chi4, D. Karasik5, J. Brain6, DP. Kiel4, K. Gautvik6. 1) Hebrew SeniorLife Inst Aging Res and Harvard Medical School, Boston, MA; 2) Molecular and Integrative Physiological Sciences Program, Harvard School of Public Health, Boston, MA; 3) Oslo University Hospital, Ulleval, Norway; 4) Medical Informatics Center, Harvard Medical School, Boston, MA; 5) Dept. Environmental Health, Harvard School of Public Health, Boston, MA; 6) Department of Medical Biochemistry, Oslo University Hospital, Ulleval, Norway.

Metabolomics is the quantitative measurement of dynamic metabolic response to pathophysiologic stimuli or genetic modification making metabolome quintessence of biological processes. To identify metabolite signatures for osteoporotic fractures and to understand its underlying biological implication, we measured ~500 metabolites in bone tissue and serum from osteoporotic hip fracture patients (OP) and age-matched normal-bone mineral density controls. Bone tissues were obtained from trans-iliac biopsies in 84 post-menopausal Caucasian women. Transcriptome profiling of bone biopsies was measured by Affymetrix microchips. GWAS genotyping was done in all samples to identify variants associated with metabolites and gene expression. Metabolites were measured by three types of mass spectrometers (MS). MS signatures of metabolites were matched to a small-molecule library. Age, BMI, estrogen use, cigarette smoking, menopause age and medication were collected and used in statistical analyses. To find metabolite signature, we applied a support vector machine algorithm with a nonlinear classification and identified 12 endogenous metabolites significantly differed between OP and controls. These 12 metabolites are involved in fatty acid amidases; lipid oxidation; bile acid metabolism; tryptophan metabolism and redox homeostasis. Among them, significant accumulation of 5 dicarboxylic fatty acids suggests potentially suppressed mitochondrial β-oxidation in OP. Higher levels of the cannabinoid (CB) receptors agonist, oleamide, were observed in OP, indicating a disruption of CB signaling. Serotonin (5HT) was also higher in OP. To further characterize underlying mechanisms, we performed regression analysis between transcriptome and these 12 metabolites, adjusted for covariates, sex, and age. The regression of 71 genes was associated with these 12 metabolites at p < 5x10^-6. AR, FGF2, PRKAA2 and several G-protein coupled receptors expressions associated with oleamide, providing supportive evidence of the involvement of CB receptors in OP. A pathway analysis pointed out three distinct cellular functional pathways. In summary, we identified a ‘signature’ of 12 endogenous metabolites associated with osteoporosis in postmenopausal Caucasian women. We used the transcriptome to further characterize their functional implications. A replication study of metabolite signatures in ~1,000 Framingham Study participants is underway to further validate these findings.
1049F  
Hypothesis independent pathway analysis identifies biologic pathways influencing susceptibility to glaucoma. J.L. Wiggs1, J.N. Cooke Bailey2, L.R. Pasquale3, S.J. Loomis4, J.H. Kang5, B. Yasan6, M. Brilliant7, W. Christen8, J.H. Finger9, D. Gasterland10, T. Gaasterland10, R.K. Lee10, R.P. Lichte10, Y. Lu11, S.E. Moroi12, L.M. Olson12, J.E. Richards12, J.S. Schuman12, W.K. Scott12, K. Singh12, A. Sit12, D. Vollrath12, G. Wolfsstein12, D.J. Zock12, K. Zhang12, M. Allingham12, M.A. Pericak-Vance12, R.N. Weinreb18, M.A. Hauser19, J.J. Haines19. NEIGHBORHOOD Consortium, 1) Dept Ophthalmology, Harvard Med Sch, MEEI, Boston, MA; 2) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Ophthalmology and Anatomy/Cell Biology, University of Iowa, Coll of Medicine, Iowa City, IA; 4) Eye Doctors of Washington, Chevy Chase, MD; 5) Scripps Genome Center, University of California at San Diego, San Diego, CA; 6) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 10) Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, MI; 11) Department of Ophthalmology, Duke University Medical Center, Durham, NC; 12) Department of Ophthalmology, UPMC Eye Center, University of Pittsburgh, Pittsburgh, PA; 13) Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 14) Department of Ophthalmology, Stanford University, Palo Alto, CA; 15) Department of Ophthalmology, Mayo Clinic, Rochester, MN; 16) Department of Genetics, Stanford University, Palo Alto, CA; 17) Wilmer Eye Institute, Johns Hopkins University Hospital, Baltimore, MD; 18) Department of Ophthalmology, University of California, San Francisco, CA; 19) Department of Medicine, Duke University Medical Center, Durham, NC.

Using genome-wide SNP data, we assessed biologic pathways as annotated in the KEGG database for association with risk of primary open-angle glaucoma (POAG) overall, as well as within the high tension POAG subgroup with increased susceptibility to optic nerve degeneration. Our pathway analysis included all SNPs in the GWAS to identify meaningful associations that did not meet the genome-wide significance threshold. The pQUANT pathway analysis (by Random-Structure) algorithm used in this study creates a null distribution of random pathways to mimic the size and structure of the actual pathways. The algorithm uses this null distribution to determine whether statistical associations are due to the biologic properties of the pathway, or instead, its size and structure. We conducted a POAG case-control analysis of 3108 cases and 3430 controls from the GLAUGEN and NEIGHBOR studies genotyped on the Illumina 660W quad platform. We first analyzed each study individually for single-SNP associations using logistic regression models to test for association. A subsequent meta-analysis of the two studies was performed. After multiple testing correction, we found 14 pathways associated with risk of POAG overall (permuted p<0.001 for all pathways). These pathways were loosely grouped into three categories: metabolism, cellular adhesion and signaling, and autoimmune disorders. Twelve of these pathways contained the same HLA genes and after removal of these genes, 11 of the pathways were no longer significant, while the HLA gene set was highly significant (p<0.001). In the normal tension subgroup analysis 7 pathways were associated with disease risk (p<0.001) falling into 5 general categories: metabolism (butanoate and ketone bodies), amino acid synthesis, MAPK signaling; hedgehog signaling and glycosaminoglycan biosynthesis. Only the butanoate metabolism pathway was significant for both POAG overall and the normal-tension subgroup. The butanoate pathway includes enzymes responsible for GABA metabolism, which can influence ganglion cell function in the retina. Overall, these data provide insight into the complex genetic etiology of glaucoma-related optic nerve degeneration and provide hypotheses for future functional studies.

1050W  

BACKGROUND: Meniscal tears predispose individuals to the development of osteoarthritis (OA). This is due in part to the mechanical changes that occur within the knee joint following meniscal injury. Additionally, molecular differences between arthritic and normal knee joints have been shown. It is unclear whether these molecular changes start to develop. We hypothesize that, in addition to the mechanical changes important molecular changes occur within the knee joint soon after the meniscal tear contributing to eventual development of OA. The purpose of this study is to characterize the gene expression profile of synovial fluid (SF) following a meniscal tear; looking for early expression of inflammatory and arthritis-related genetic markers. METHODS: SF was collected from 11 patients (<48yrs) undergoing an arthroscopic partial meniscectomy at the time of surgery. RNA was extracted from cell pellets precipitated from the SF. Gene expression analysis was completed using Illumina HumanHT-12 array and Lumi R code (Biorconductor) software. A heatmap identified patterns among gene expression profiles and pathway analysis was performed by Metacore software. T-tests with Bonferroni correction compared subsets of genes between individuals with recent injury (<2 months; short duration) versus longer time lapse (long duration) until meniscectomy. RESULTS: The top 5% genes expressed across all samples mapped to inflammatory and cytoskeleton remodeling pathways. Included in the most abundant genes; IL1R1, TSPAN16, CXCL9, IL18 and IL18R1 previously. Clustering of 3 samples with short duration was observed from the heat map and when compared to those samples with long duration across a subset of 340 genes with previous links to inflammatory or arthritic conditions, the following showed significant differential expression (p < 1.47x10^-4); IL1A, IL1F2, NFKB1, IL10RB, IL18, ILF3, IL13RA, BMP2K and IL10. These data support the inception of changes early in the disease process as well as indicated continued changes in expression profiles over time after injury. CONCLUSION: Our results support the hypothesis that molecular changes predisposing meniscal injury patients to OA may occur earlier after injury than previously reported. These findings suggest the possibility for early interventional treatments to halt these molecular changes and reduce the susceptibility of patients to OA following a meniscal tear.

1051T  
Genetic association study of adaptor protein complex 4 with cerebral palsy in Han Chinese population. Q. Xing1,2, H. Wang1,2, T. Li3, M. Chen1,2, Q. Shang4, D. Zhu4, L. Wang5,6, Q. Li5,6, L. He5,6, C. Zhu5,6. 1) Children’s Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai, China; 2) Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University; Shanghai, China; 3) Department of Pediatrics, the Third Affiliated Hospital of Zhengzhou University; Zhengzhou, China; 4) Department of Pediatrics, Zhengzhou Children’s hospital; Zhengzhou, China; 5) Center for Brain Repair and Rehabilitation, University of Gothenburg; Sweden.

Cerebral palsy (CP) is caused by injury or developmental disturbances to the immature brain and leads to substantial motor impairments. Adaptor protein complex 4 (AP-4) plays a key role in vesicle formation, trafficking, and sorting processes, which are critical for brain development and function. AP-4 consists of four subunits encoded by AP4E1, AP4B1, AP4M1 and AP4S1 gene respectively. AP-4 is identified to involve in trafficking of the a-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) glutamate receptor. A number of studies have pointed to the involvement of AP-4-mediated vesicular trafficking pathway in the etiology of cerebral palsy, the most notable of which are the causative mutations identified in each of them that have recently been reported in different CP families. We therefore postulated that variations in AP-4 coding genes may exert an important role in the susceptibility to CP. In the present study, sixteen SNPs were gene typed among 517 CP patients and 502 healthy controls from the Chinese Han population. We systematically analyzed the association of the AP4E1, AP4B1, AP4M1 and AP4S1 genes with CP on the basis of clinical characters. No significant associations were found between these variants and the risk of overall CP. Subgroup analysis showed that rs1217401 of AP4E1 was significantly associated with CP as a sequela of hypoxic-ischemic encephalopathy (HIE) (HIE + CP) (allele: p = 0.042151; genotype: p = 4.46x10^-6). The current results indicate that 16 variants studied in four units of AP-4 have no detected effects on the susceptibility to overall CP, but AP4B1 is a susceptible gene for HIE + CP in the Chinese population.
1052F Association analysis of candidate gene polymorphisms in Asthma, Rhinitis and Chronic Bronchitis: preliminary results from the GEIRD study. C. Bombieri1, F. Belpinati1, A. Balitan1, A.R. Lo Presti1, G. Malerba1, S. Accordini2, L. Calciano2, M. Ferrari2, I. Perbellini3, P.F. Pignatti1, R. De Marco2. 1) Dpt. of Life and Reproduction Sciences, Section of Biology and Genetics, University of Verona, Verona, Italy; 2) Dpt. of Medicine and Public Health, Section of Epidemiology and Medical Statistics, University of Verona, Verona, Italy; 3) Unit of Internal Medicine, University of Verona, Verona, Italy; 4) Unit of Occupational Medicine, University of Verona, Verona, Italy. Within The Gene Environment Interactions in Respiratory Diseases (GEIRD) study (Int Allergy Immunol 2010;152:255-263), we performed an association analysis, in a large and accurately defined series of Italian subjects, to investigate the genetic involvement in the susceptibility to respiratory diseases. The study population included 1000 subjects (aged 20-86 years) from the general population, enrolled between 2007 and 2010, through a two-stage screening process (postal questionnaire and clinical examination). Cases and controls were diagnosed on the basis of the answers to a detailed interview, to collect informations about family history for respiratory diseases, individual and ecological exposures, diet, smoking habits, medication, life-styles, and of pre/post bronchodilator spirometry, methacholine challenge, and skin prick tests. According to collected data, subjects were hierarchically classified as follows: 342 asthma, 7 COPD (with no asthma), 112 chronic bronchitis (with neither asthma nor COPD), 205 rhinitis (without asthma, COPD or chronic bronchitis) patients, and 312 controls (subjects without respiratory diseases). Classification was not possible for 22 subjects which did not completely meet diagnostic criteria of neither cases nor controls. A total of 971 subjects were genotyped, by a customized version of the GoldenGate Genotyping assay (Illumina), for a panel of 394 SNPs (Haplotype-Tagging-SNPs), representative of 63 candidate genes with a previous indication of possible association to the studied diseases. COPD subjects (due to the limited number) and unclassified subjects were not considered in the present analysis. A preliminary association analysis based on allele frequency comparison was performed. Presence of association (unadjusted p<0.001) was observed between polymorphisms of RAD50-IL13-LR4 region and past-asthma, PDE4D and ever asthma with atopy, IL13-IL4 region and past-asthma, PDE4D and ever asthma with atopy, SPINK5 and rhinitis, GPR154 and chronic bronchitis with atopy, PDE4D and chronic bronchitis with atopy, and PDE4D and chronic bronchitis with atopy. The past-asthma analyses have predominantly included diverse European and Asian populations. The present study now investigated a large and phenotypically well-characterized sample of Arabian origin from the Yemen population. Individuals included nsOCP patients (n=310), their relatives (n=107) and unaffected controls (n=423). Patients were recruited within the context of surgical outreach programs that took place in Yemen between 2010 and 2012. Controls were recruited from blood donors of the same ethnic origin in the same area. Blood samples of patients/controls were transferred to the Institute of Human Genetics, Bonn where DNA was extracted. As research into nsCL/P has seen major advances by the recent application of genome-wide association studies (GWAS), we analyzed the genome-wide significant variants that have been previously been identified in 1) European, 2) Middle Eastern and Asian samples and in the nsCPO group (n=45) of the Yemeni patients. Twenty-four markers (SNPs, single-nucleotide polymorphisms) representing fifteen loci were genotyped using MassArray spectroscopy (n=23) or TaqMan assay (n=1). Loci included the European high-risk locus at 8q24.21 (rs987525) and the IRF6 locus. Statistical analyses using Armitage trend test revealed SNPs at nine of these loci to show nominal significant P-values in nsCL/P, the lowest P-value observed being 2.46×10^-22 for rs987525, with an relative risk of 1.80 (95% confidence interval: 1.27-2.55) for the heterozygous and 2.45 (95% CI: 1.54-3.90) for the homozygous genotype. In total, markers at three of the loci (8q24.21, 9q22 and 13q31) withstood correction for multiple testing. In the nsCPO group, no nominal significant associations were identified. This observation is consistent with results from other populations and might reflect the fact that there is little genetic overlap between nsCL/P and nsCPO. Our results reveal that a majority of nsCL/P risk loci identified in European and Asian ethnicities are also conferring risk for nsCL/P in the Arabian population.

1053W Extending the population spectrum for nonsyndromic orofacial clefting: Recruitment and genetic analyses in an Arabian population from Yemen. K.U. Ludwig1,2, K. Afdal1, A.C. Boehler1,2, B. Lippke2, N. Daratsianos3, M.M. Noethen1,2, A. Jaeger4, M. Knapp5, E. Mangold1. 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life and Brain Center, Bonn, Germany; 3) Faculty of Dentistry, University of Science and Technology, Sanaa, Yemen; 4) Department of Orthodontics, University of Bonn, Bonn, Germany; 5) Institute of Medical Biometry Informatics and Epidemiology, University of Bonn, Germany. Nonsyndromic orofacial clefting (nsOFC) is a frequent congenital disorder, and has a genetically complex etiology. There are two common subtypes of nsOFC, namely nonsyndromic cleft lip with or without cleft palate (nsCL/P) and nonsyndromic cleft palate only (nsCPO). Research into the genetics of clefting is ongoing. nsCPO has identified numerous genetic susceptibility loci. The past genetic analyses have predominantly included diverse European and Asian populations. The present study now investigated a large and phenotypically well-characterized sample of Arabian origin from the Yemen population. Individuals included nsOCP patients (n=310), their relatives (n=107) and unaffected controls (n=423). Patients were recruited within the context of surgical outreach programs that took place in Yemen between 2010 and 2012. Controls were recruited from blood donors of the same ethnic origin in the same area. Blood samples of patients/controls were transferred to the Institute of Human Genetics, Bonn where DNA was extracted. As research into nsCL/P has seen major advances by the recent application of genome-wide association studies (GWAS), we analyzed the genome-wide significant variants that have been previously been identified in 1) European, 2) Middle Eastern and Asian samples and in the nsCPO group (n=45) of the Yemeni patients. Twenty-four markers (SNPs, single-nucleotide polymorphisms) representing fifteen loci were genotyped using MassArray spectroscopy (n=23) or TaqMan assay (n=1). Loci included the European high-risk locus at 8q24.21 (rs987525) and the IRF6 locus. Statistical analyses using Armitage trend test revealed SNPs at nine of these loci to show nominal significant P-values in nsCL/P, the lowest P-value observed being 2.46×10^-22 for rs987525, with an relative risk of 1.80 (95% confidence interval: 1.27-2.55) for the heterozygous and 2.45 (95% CI: 1.54-3.90) for the homozygous genotype. In total, markers at three of the loci (8q24.21, 9q22 and 13q31) withstood correction for multiple testing. In the nsCPO group, no nominal significant associations were identified. This observation is consistent with results from other populations and might reflect the fact that there is little genetic overlap between nsCL/P and nsCPO. Our results reveal that a majority of nsCL/P risk loci identified in European and Asian ethnicities are also conferring risk for nsCL/P in the Arabian population.

1054T Further evidence suggesting a role for variation in ARHGAP29 in nonsyndromic cleft lip/palate. L. Maill1, A. Leitra1,2, J.B. Mulliken3, S. Stiller3, S.H. Blanton5, J.T. Hecht1,2, 1) Department of Pediatrics, Pediatric Research Center, University of Texas Health Science Center Medical School at Houston, Houston, TX; 2) Department of Endodontics, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 3) Center for Craniofacial Research, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 4) Children’s Hospital, Boston, MA; 5) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL. Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a common birth defect occurring in approximately 1 in 700 births. The complex etiology of NSCL/P reflects multiple genetic and environmental factors acting individually or in concert. A variety of research approaches, including candidate gene, genome-wide linkage, and genome-wide association studies (GWAS), have been used to identify the etiologic genes contributing to NSCL/P. While numerous genes have been implicated, only a few have been replicated across datasets. ARHGAP29 was suggested as a candidate gene for NSCL/P by being located in close proximity to ABCA4 (1p22), a gene previously identified in a GWAS of NSCL/P. Rare, potentially damaging, coding variants in ARHGAP29 were found in NSCL/P cases, and its expression was detected during craniofacial development in mice. Taken together, these findings suggest a role for this gene in NSCL/P. In this study, we investigated if variations in ARHGAP29 were associated with NSCL/P. Five SNPs flanking and within ARHGAP29 were genotyped in our NSCL/P datasets consisting of simplex and multiplex families of white NonHispanic (WHN, primarily western European) and Hispanic race/ethnicity. Family-based association tests were performed, stratified by ethnicity and family history of NSCL/P. P-values ≤ 0.05 were considered statistically significant. Results showed strong association of three ARHGAP29 SNPs with NSCL/P in the WHN families. Two intronic SNPs (rs1541098 and rs3789688) showed association with NSCL/P in all WHN families (p=0.0005 and p=0.0002, respectively), and in the simplex WHN families. A SNP in the 3’ UTR (rs1576593) also showed an association with NSCL/P, in all WHN families (p=0.002) and the multiplex subset (p=0.002). No associations were found between variants in ARHGAP29 and NSCL/P in the Hispanic datasets. This study further supports ARHGAP29 as a candidate gene for human NSCL/P in families of Caucasian descent.
Association of GABRG2 rs211307 polymorphism with susceptibility to epilepsy in Asians: a multicentre case control study and meta-analysis. B.S. Haerian1, L. Baum2, P. Kwan3, S.S. Cherney4, H.J. Tan5, A.A. Raymond6, Z. Mohamed1. 1) Pharmacology, University of Malaya, Kuala Lumpur, Malaysia; 2) School of Pharmacy, The Chinese University of Hong Kong, Shatin, Hong Kong, China; 3) Division of Neurology, Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, Hong Kong, China; 4) Department of Psychiatry and The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China; 5) Department of Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia.

The gamma-aminobutyric acid receptor subunit gamma-2 (GABRG2) gene encodes GABR-gamma2 protein, which has been implicated in susceptibility to epilepsy. Several studies have investigated whether the synonymous GABRG2 rs211307 polymorphism is a risk factor for various epilepsy types including febrile seizure (FS), idiopathic generalized epilepsy (IGE) and symptomatic epilepsy (SE), however results have been inconclusive. Therefore, we examined association of this polymorphism in FS, IGE, SE, and cryptogenic epilepsy (CE), through both a multicenter case control study and a meta-analysis. rs211307 was genotyped in Hong Kong Chinese and in Malaysian Chinese, Indian, and Malay participants. Genotypes of 5101 participants, of which 1769 were from Malaysia and 3332 from Hong Kong, were included in this case-control study. Of the 1179 patients, 66 percent; were Chinese, Indians, and Malays, respectively. Of the Chinese patients, 37 percent; were from Malaysia and 63 percent; were from Hong Kong. Significant association was observed between rs211307 polymorphism and susceptibility to SE (T vs. C p=0.02 and CT vs. CC, p=0.02) in overall Chinese and to IGE in Chinese from Malaysia (TT vs. CC, p=0.01), but not to CE. Meta-analysis revealed a strong association between rs211307 with FS and SE in Asians for alleles (p=0.02, p=0.002 and p<0.00001, respectively) and for all genotype models. Our data suggests that GABRG2 rs211307 polymorphism is a risk factor for susceptibility to SE and FS in Asians, particularly in Chinese.
1057T
APOEe2 homozygous individuals are underrepresented among elderly Brazilian population, M.S. Naslavsky1, M.L. Lebrão2, Y.A.O. Duarte2, E. Amano2, Laval University, Quebec, Canada; 3. A.B. Mendes3, A.S. Rodrigues3, D. Bandeira1, D. Schlesinger5, L.T. Grinnberg3,6, C.K. Suwomo3, R.E.P. Leite3, R.E.L. Ferretti3, C.A. Pasqualucci3, J.M. Farfès3, R. Nitrini3, W. Jacob Filho3, M. Zatz1, 1. Human Genome and Stem Cell Center, University of Sao Paulo (USP), Sao Paulo, SP, Brazil; 2. Public Health Faculty, USP, Sao Paulo, SP, Brazil; 3. Medical School, USP, Sao Paulo, SP, Brazil; 4. Instituto do Cerebro, Hospital Israelita Albert Einstein, Sao Paulo, SP, Brazil; 5. Mendelcis Anisele Genomica, Sao Paulo, SP, Brazil; 6. Department of Neurology, University of California, San Francisco, CA, USA.

The genome variation of healthy individuals undergoing aging has been the focus of worldwide interest and research. In order to enhance our understanding on human genetic variability our group is building a comprehensive collection of data from samples of elderly individuals from Sao Paulo (Brazil) which has a mixed population of about 12 million inhabitants characterized by a great ethnic variability. Up to now, the cohort is composed by more than 1400 individuals who will be submitted to whole genome analysis. However, a preliminary analysis of apolipoprotein E, a key protein in the metabolism and transport of lipoproteins and cholesterol and coded by chromosomal region 19, has shown unexpected results and called our attention. APOE variants have been highly associated with the physiopathology of late onset Alzheimer’s disease (LOAD) and thus genotyping the different alleles becomes a standard procedure within studies of aging-related cognitive impairment. The main isoforms are APOE e2, APOE e3 and APOE e4. While APOE e4 is thought to increase AD risk, APOE e2 has been considered to be a protective variant. DNA samples from 1447 individuals, 60 years old or older were analyzed including 523 arteriosclerotic cases. Up to now, 1000 APOE genotypes were obtained by allele-specific amplification. Genotypic and allelic frequencies were derived by direct count. Surprisingly, although allelic frequencies were within the expected frequency according to other population studies, the APOE e2 homozygote was underrepresented and shifted from Hardy-Weinberg Equilibrium (HWE). The same analysis in a subset of 519 samples of individuals deceased from natural causes (aged 50 to 102) obtained from the largest brain bank of Sao Paulo showed no deviation from HWE. It is suggested that even considering APOE e2 an LOAD protective allele, carriers of APOE e2 isoform may be at risk for early vascular disorders such as hyperlipoproteinemia. We hypothesize that early vascular disorders may increase the mortality rate of APOE e2 homozygotes and, therefore, be responsible for the decreased frequency of individuals homozygous among elderly subjects. We are currently increasing the sample size and analyzing the correlation between vascular causes of mortality and expected APOE e2 homozygous frequency. We believe this may bring important contribution to our understanding on the role of APOE e2 polymorphism in LOAD and vascular disorders suscepti- bility.

1059W

Abnormal lipid concentrations are risk factors for atherosclerosis and cardiovasculardisease. Each constitutional type classified by Sasang constitutioinal medicine, a Korean constitutional medicine, has been shown to have a distinctive pathological characteristic for cardiovascular disease risk. Here, we performed an association analysis between lipid-related traits and genetic variants from several genome-wide association studies according to Sasang constitutional types classified using an integrated diagnostic model consisting of face, body shape, voice, and questionnaire information. We estimated the associations of 26 lipid-associated variants in 20 loci using multiple regression analyses in 2 Korean populations, i.e., 8,597 individuals. Of the 26 variants, 12 were significantly associated with lipid levels, including LDL cholesterol, HDL cholesterol, and triglycerides. In subgroup analysis of Tae-Eum (TE) and non-Tae-Eum (NTE) types (each 2,664 individuals) on the basis of tertiles of probability values from the diagnostic model for TE type harboring predominant cardiovascular risk, the associations of 3 variants near APOA5-APOA4-APOC3-APOA1 (rs6589566 on triglyceride: p = 8.90×10-11, APOE-APOC1-APOC4 (rs4206383 on triglyceride: p = 1.63×10-5, and LIPG (rs2156552 on HDL cholesterol: p = 0.00428) remained significant in the TE type, while those of 3 variants near ANGPTL3 (rs10889533 on triglyceride: p = 0.00233), APOA5-APOA4-APOC3-APOA1 (rs6589566 on triglyceride and HDL cholesterol: p = 0.0096 and 0.00055, respectively), and LPL (rs6586891 on HDL cholesterol: p = 0.00020) remained significant in NT type. Interestingly, the minor allele effects of the lipid-associated variants in TE type had a harmful influence on lipid risk, whereas those of the variants in NT type had neutral influences due to a compensating effect among the associated variants. These findings supported that the genetic susceptibility to lipid risk is affected by the minor allele effect of the variants, which may predispose TE type subjects to a high cardiovascular disease risk.

1059T
Pleiotropic effects of three SLE associated functional variants within IFIH1 linked to several autoimmune diseases. J.E. Molineros, C.A. Pasqualucci, T.A.B. Mendes, J.-C. Bérubé, M.S. Naslavsky, Y. Bossé, 1) Centre de recherche Institut universitaire de cardiologie et de pneumologie de Quebec, Quebec, Canada; 2) Department of Molecular Medicine, Laval University, Quebec, Canada; 3) Department of Surgery, Université Laval, Quebec, Canada; 4) Department of Medicine, Laval University, Quebec, Canada; 5) Department of Molecular Medicine, Université Laval, Quebec, Canada; 6) Department of Genetics, University of Sao Paulo, Sao Paulo, SP, Brazil.

Systemic lupus erythematosus (SLE)[Molineros et al. 2013, PLos Genet 9, e1003222]. These three variants are rs1990760 (A690T), rs10930046 (H460R) and intronic SNP rs13023380. IFIH1 has been related to viral response and apoptosis functions. There is a close correlation between viral response and the development of autoimmune disease (AD). Therefore, ADs are likely to share susceptibility variants/genes, especially those related to viral immune response. In fact, the pleiotropic effect has been reported for IFIH1 rs1990760 variants rs1990760 with some ADs. Objective The objective was to extend the analysis of pleiotropic effects to six ADs and 2 non-ADs, and to identify other possible novel AD associated variants within IFIH1. Approach We performed an imputation based association analysis on IFIH1 with 1000 Genomes Project variants using MACH. Only SNPs with imputation quality >0.8 and minor allele frequency >0.05 were considered. Results: Five variables were selected from the factor analysis. From these, three clinical sub-groups were produced: smoking history, low atopy and low lung function, high atopy, and young non-smoking with average atopy. Genotype frequencies of four SNPs were significantly different between sub-groups (p-value<0.05). One is located near IKZF4 and the others within RCBTB1, DPP10 and IL33. The frequency of the asthma risk allele for SNP on DPP10 (rs1435879) was less frequent in the highly allergic sub-group compared to the other sub-groups.

Conclusion: The genotype frequencies of four SNPs were different across homogeneous sub-groups of asthmatic patients. DPP10-rs1435879 was particularly different in a sub-group characterized as highly allergic. Whether SNPs in DPP10 can be used to modify the current asthma classification remains to be validated. This data is needed as we have evidence that there is no evidence that IFIH1 is differentially distributed within asthma cohorts in order to improve current classification algorithm, and ultimately offer targeted therapy.

1059F
GWAS-nominated variants in homogeneous asthma sub-phenotypes. E. Laviole-Charland1, J.-C. Bérubé1, M. Laviole1, L.-P. Boulet1, Y. Bossé2, 1) Centre de recherche Institut universitaire de cardiologie et de pneumologie de Quebec, Quebec, Canada; 2) Department of Molecular Medicine, Laval University, Quebec, Canada.

Purpose: Asthma is a heterogeneous disease and subgrouping patients with similar clinical characteristics is likely to be important to improve our molecular understanding of this disease. Large-scale genome-wide association studies (GWAS) have identified many single nucleotide polymorphisms (SNPs) robustly associated with asthma. The goal of this study is to test GWAS-nominated SNPs in phenotypically similar asthmatic patients.

Methods: The number of clinical variables was reduced by factor analysis. With selected variables, k-mes clustering was applied on 523 asthma patients to produce more homogenous subgroups of patients with asthma. Genotypes from 49 independent SNPs selected from previous GWAS on asthma were then compared between these phenotypically similar subgroups.

Results: Five variables were selected from the factor analysis. From these, three clinical sub-groups were produced: smoking history, low atopy and low lung function, high atopy, and young non-smoking with average atopy. Genotype frequencies of four SNPs were significantly different between sub-groups (p-value<0.05). One is located near IKZF4 and the others within RCBTB1, DPP10 and IL33. The frequency of the asthma risk allele for SNP on DPP10 (rs1435879) was less frequent in the highly allergic sub-group compared to the other sub-groups.

Conclusion: The genotype frequencies of four SNPs were different across homogeneous sub-groups of asthmatic patients. DPP10-rs1435879 was particularly different in a subgroup characterized as highly allergic. Whether SNPs in DPP10 can be used to modify the current asthma classification remains to be validated. This data is needed as we have evidence that there is no evidence that IFIH1 is differentially distributed within asthma cohorts in order to improve current classification algorithm, and ultimately offer targeted therapy.
2061F

Are genes previously associated with schizophrenia also predictive of dimension-specific psychotic experiences in adolescence? D. Sieradzka1, R.A. Power2, F. Dudbridge3, E.L. Meaburn1, R. Plomin1, A. Ronald1
1) Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck, University of London, London, UK; 2) MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London, UK; 3) Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK.

Evidence suggests that a considerable amount of variance in schizophrenia liability can be accounted for by common genetic variants. As psychosis can be viewed as a continuum, here we test the hypothesis that these common variants associated with schizophrenia also influence sub-clinical psychotic experiences in adolescence. Our aim was to test whether polygenic risk scores from schizophrenia GWAS and specific single nucleotide polymorphisms (SNPs) previously identified as risk variants for schizophrenia (Bergen & Petryshyn, 2012), were associated with dimension-specific psychotic experiences. Six dimension-specific psychotic experiences (self-reported: paranoia, hallucinations, cognitive disorganisation, grandiosity, anhedonia, and parent-rated negative symptoms), as measured by the Specific Psychotic Experiences Questionnaire (SPEQ), were assessed in a community sample of 2,130 16-year-olds. First, polygenic risk scores were calculated using estimates of the logs of odds ratios from the Psychiatric Genomics Consortium (PGC) GWAS mega-analysis of schizophrenia and polygenic risk analyses were conducted. Schizophrenia polygenic risk scores were a significant predictor of anhedonia at pT=.40, pT=.50 and pT=.60 (all p-values<.008; corrected) and parent-rated negative symptoms at pT=.01 only (p-value=.004; corrected); however, not in the expected direction. A second, individual SNP analyses were performed to test for associations between dimension-specific psychotic experiences and 28 SNPs previously associated with schizophrenia. A significant association was found between the SPEQ paranoia dimension and rs17512836 (TCF4; p=.02, post corrected). This study presents the first empirical test of whether aggregated common variants and single SNPs associated with schizophrenia are also predictive of dimension-specific psychotic experiences in adolescence. The results do not provide support for the hypothesis but due to the limited power for identifying very small genetic effects, further research in larger samples sizes is required.

1062W

Association of GALNT10 genetic variants with adiposity in African Americans. M.E. Stromberg1,2, J.M. Hester3,2, P. Mudgal1,2, J. Li1,2, P.J. Hicks1,2, B.L. Freedman1,2, D.W. Bowden2,4,5, M.C.Y. Ng1,2,1
1) Center for Diabetes Research, Wake Forest University School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, NC; 3) School of Medicine, Wake Forest University School of Medicine, Winston-Salem, NC; 4) Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC; 5) Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC.

Despite the higher prevalence of obesity (35.7%) in African Americans (AAs) compared to non-Hispanic whites (23.7%), genetic contributors to obesity in AAs remain poorly understood. Recent genome-wide association studies (GWAS) identified a region on 10q26.32 of the GALNT10 associated with body mass index (BMI) in AAs. GALNT10 is a member of the N-acetylgalactosaminyltransferase family and is located in the Golgi apparatus. The 20 known members of this family have been reported to have varying expression patterns in different tissues and multiple associations with metabolic traits. GALNT10 is expressed in the hypothalamus, which regulates hunger. We examined multiple SNPs in the associated linkage disequilibrium (LD) region for association with BMI and other adiposity measures including waist circumference to hip ratio (WHR), visceral adipose tissue (VAT), subcutaneous adipose tissue (SAT) and VAT:SAT ratio (VSR). The 9 common variants tagging the LD block surrounding previous GWAS top hits and 12 exonic variants from the NHLBI ESP GO database were genotyped. Coding SNPs were prioritized for minor allele frequency (MAF) ≤0.05 with predicted functional impact using PolyPhen. Variants were prioritized in association with BMI, WAIST, WHR, VAT, SAT and VSR in a meta-analysis of African American cohorts totalling up to 4,992 subjects. The most significantly BMI-associated SNPs were common (MAF between 0.32-0.44, noncoding). The minor allele was associated with higher BMI (p=3.64E-06), followed by rs7708584 (p=3.71E-05), rs815611 (p=4.65E-06), rs7719067 (p=6.99E-06), rs2033195 (p=7.26E-06), rs1366219 (p=9.91E-06) and rs4958361 (p=0.001). Associations were seen for BMI, not other adiposity traits. There was minimal evidence for association of coding or rare variants (individually or jointly with sequence kernel association test analysis). Common variants near GALNT10 are associated with BMI (not other body size), but not other measures of adiposity, and coding variants in the gene do not contribute measurably to the association. Fine mapping of the GALNT10 gene region will assist in detecting causal variants for BMI in the region.

1063T

Investigation of the rs2157719 SNP in the CDKN2B-AS1 gene in a Primary Open-Angle Glaucoma Brazilian population. H.F. Nunes1, J.P.C. Vasconcellos2, V.P. Costa2, N.I.T. Zanchin2, M.B. Melo2
1) CBMEG, University of Campinas, Campinas - SP, Brazil; 2) Department of Ophthalmology, Faculty of Medical Sciences, University of Campinas, Campinas - SP, Brazil; 3) Fiocruz, Curitiba - PR, Brazil.

Primary open-angle glaucoma (POAG) is a chronic neurodegenerative disease that leads to progressive damage of retinal ganglion cells resulting in visual field loss. Glaucoma is recognized as the main cause of irreversible blindness worldwide. Although the pathophysiology of glaucoma is not well understood, positive family history is one of the most important risk factors for glaucoma development. Single nucleotide polymorphisms (SNPs) have been widely associated with glaucoma, although the frequencies vary among different populations. CDKN2B-AS1 gene encodes a long noncoding RNA, and it is likely to play a role in regulating the expression of genes at the 9p21 locus through epigenetic mechanisms. CDKN2A encodes 2 proteins, p16INK4A and p14ARF, which is a member of the N-acetylgalactosaminyltransferase family and is expressed in the hypothalamus, which regulates hunger. We examined multiple SNPs in the associated linkage disequilibrium (LD) region for association with BMI and other adiposity measures including waist circumference to hip ratio (WHR), visceral adipose tissue (VAT), subcutaneous adipose tissue (SAT) and VAT:SAT ratio (VSR). The 9 common variants tagging the LD block surrounding previous GWAS top hits and 12 exonic variants from the NHLBI ESP GO database were genotyped. Coding SNPs were prioritized for minor allele frequency (MAF) ≤0.05 with predicted functional impact using PolyPhen. Variants were prioritized in association with BMI, WAIST, WHR, VAT, SAT and VSR in a meta-analysis of African American cohorts totalling up to 4,992 subjects. The most significantly BMI-associated SNPs were common (MAF between 0.32-0.44, noncoding). The minor allele was associated with higher BMI (p=3.64E-06), followed by rs7708584 (p=3.71E-05), rs815611 (p=4.65E-06), rs7719067 (p=6.99E-06), rs2033195 (p=7.26E-06), rs1366219 (p=9.91E-06) and rs4958361 (p=0.001). Associations were seen for BMI, not other adiposity traits. There was minimal evidence for association of coding or rare variants (individually or jointly with sequence kernel association test analysis). Common variants near GALNT10 are associated with BMI (not other body size), but not other measures of adiposity, and coding variants in the gene do not contribute measurably to the association. Fine mapping of the GALNT10 gene region will assist in detecting causal variants for BMI in the region.
**1064F**

**Genetic Testing for Age-Related Macular Degeneration in an Armenian Population.** K.W.S. Small1,2,3, A.A. Abraamyan1,2,3, B.Z. Zanke1, P.R. Ramamothry1. 1) Molecular Insight Research Foundation, Los Angeles, CA; 2) Cedars-Sinai Regenerative Medicine Institute, Los Angeles, CA; 3) Macula & Retina Institute, Los Angeles, CA; 4) Arctic Dx - Macula Risk; Bonita Springs, FL.

Purpose: Age-related macular degeneration (ARMD) has a significant genetic influence, especially in Caucasian groups. The Armenian ethnogenesis dates back at least 3,000 years and is considered to be a genetically isolated population. Additionally, the Armenian genocide created a genetic bottleneck. Because of these demographics, we hypothesized that the genetic involvement in ARMD may be different than that of other Caucasian populations. To our knowledge, this is the first reported evaluation of the genetic contribution of ARMD susceptibility genes in an Armenian population.

Methods: A retrospective review was performed of 38 Armenian patients with wet (exudative) ARMD who had genetic analysis using the commercially available Macula Risk genotyping method. We obtained buccal cheek swabs which were sent to Advanced Diagnostic Laboratories for routine clinical testing of previously documented ARMD risk alleles: Complement Factor H (CFH) with 5 Single Nucleotide Polymorphisms, Complement Component 3 (C3), Age-Related Maculopathy Susceptibility 2 (ARMS2) and mitochondrial encoded NADH dehydrogenase 2 (MT-ND2). Each genetic component adds to a specific portion of a Macula Risk score, which predicts risk of developing advanced macular degeneration. This was compared to the Macula Risk genetic database of a 766 person Caucasian population with wet ARMD. The data was analyzed using Mann-Whitney U test to show differences in age and Macula Risk score. Chi squared test with two degrees of freedom was used to show differences in the genotypes (except MT-ND2, which only required a 2x2 table). Results: The Mann-Whitney U tests showed no statistically significant difference between the Armenian and Caucasian data sets for Macula Risk score. The average age for Armenians (83.6) was higher than Caucasians (79.8), P=0.04. The Chi squared tests showed no statistically significant difference between the genotypes.

Conclusion: In Armenian and Caucasian patients with wet ARMD, there is no difference between genotypes and Macula Risk score. Although we found no statistically significant difference, the data set is relatively small and the findings are powered. Additional analysis is necessary to evaluate the reasons contributing to the age difference within these populations. In patients who are not in the ARMD risk alleles in the Armenian population compared to the Caucasian population.

**1066T**

**Positive Associations of ZIC2, RASGRF1, and SHISA6 Gene with High Myopia in Japanese.** M. Yoshikawa1,2, K. Yamashiro1, M. Oishi1,2, M. Miyake1,2, H. Nakashiri1,2, I. Nakata1,2, N. Gotoh1, Y. Kurashige1,2, K. Kumagai1, M. Oishi1, A. Tsujikawa1, R. Yamada1, F. Matsuda2, N. Yoshimura3, the Nagahama Study Group. 1) Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Center for Genomic Medicine/Inserm U.852, Kyoto University Graduate School of Medicine, Kyoto, Japan; 3) Comprehensive Human Bioscience (The Nagahama Study).

Introduction: Myopia is the most common ocular disorder in the world. High myopia, with a prevalence of 1%-2% in the general population, is distinguished from common myopia by excessive increase in axial length of the eye ball and is associated with various ocular complications leading to blindness. Recently, 24 new chromosomal loci associated with refractive error and common myopia were reported in a large scale multi-ethnic genome-wide association study (GWAS) and several loci were confirmed by another GWAS dealing myopic patients in Caucasian population. Therefore, we conducted a case-control study to investigate whether these genetic variations, which confer risk of common myopia in Caucasians, were associated with high myopia in Japanese or not. Because choroidal neovascularization (CNV) is the most common cause of visual loss related to high myopia, we also evaluated the contribution of those genetic variations to the occurrence of CNV among the high myopic patients group. Methods: 5 single nucleotide polymorphisms (SNPs) from 5 candidate genes: TOX, RDH5, ZIC2, RASGRF1 and SHISA6 reported in both of the two previous GWAS were genotyped using Taqman assay in a total of 1339 unrelated highly myopic Japanese patients. As a control, the genotype data of 3248 healthy Japanese individuals were obtained from the Nagahama Prospective Genome Study. To test for association between CNV and high myopia, we evaluated the contribution of each gene to the occurrence of CNV among the high myopic patients group. Because choroidal neovascularization is a great risk for developing vision-threatening choroidal neovascularization in the later stage of AMD, it is considered to be a genetically isolated population. Additionally, the Armenian genocide created a genetic bottleneck. Because of these demographics, we hypothesized that the genetic involvement in ARMD may be different than that of other Caucasian populations. To our knowledge, this is the first reported evaluation of the genetic contribution of ARMD susceptibility genes in an Armenian population.

Methods: A retrospective review was performed of 38 Armenian patients with wet (exudative) ARMD who had genetic analysis using the commercially available Macula Risk genotyping method. We obtained buccal cheek swabs which were sent to Advanced Diagnostic Laboratories for routine clinical testing of previously documented ARMD risk alleles: Complement Factor H (CFH) with 5 Single Nucleotide Polymorphisms, Complement Component 3 (C3), Age-Related Maculopathy Susceptibility 2 (ARMS2) and mitochondrial encoded NADH dehydrogenase 2 (MT-ND2). Each genetic component adds to a specific portion of a Macula Risk score, which predicts risk of developing advanced macular degeneration. This was compared to the Macula Risk genetic database of a 766 person Caucasian population with wet ARMD. The data was analyzed using Mann-Whitney U test to show differences in age and Macula Risk score. Chi squared test with two degrees of freedom was used to show differences in the genotypes (except MT-ND2, which only required a 2x2 table). Results: The Mann-Whitney U tests showed no statistically significant difference between the Armenian and Caucasian data sets for Macula Risk score. The average age for Armenians (83.6) was higher than Caucasians (79.8), P=0.04. The Chi squared tests showed no statistically significant difference between the genotypes. Conclusin: In Armenian and Caucasian patients with wet ARMD, there is no difference between genotypes and Macula Risk score. Although we found no statistically significant difference, the data set is relatively small and the findings are powered. Additional analysis is necessary to evaluate the reasons contributing to the age difference within these populations. In patients who are not in the ARMD risk alleles in the Armenian population compared to the Caucasian population.

**1066W**

**Associations of Age-Related Macular Degeneration Susceptibility Genes to Drusen.** K. Yamashiro1, M. Yoshikawa1, M. Miyake1,2, H. Nakashiri1,2, I. Nakata1,2, N. Gotoh1, Y. Kurashige1,2, K. Kumagai1, M. Oishi1, A. Tsujikawa1, R. Yamada1, F. Matsuda2, N. Yoshimura3, 1) Ophthalmology, Kyoto Univ Graduate Sch of Med, Kyoto, Japan; 2) Center for Genomic Medicine/Inserm U.852, Kyoto University Graduate School of Medicine, Kyoto, Japan; 3) Comprehensive Human Bioscience (The Nagahama Study).

Large drusen, typical characteristics of age-related macular degeneration (AMD), is a great risk for developing vision-threatening choroidal neovascularization in the later stage of AMD. Although susceptibility genes for AMD have been identified in previous studies, genetic background of large drusen is still unclear. We evaluated associations between large drusen and known 19 regions associated with AMD, using Japanese individuals recruited from the Nagahama Prospective Genome Cohort for the Comprehensive Human Bioscience (The Nagahama Study). Among the participants with age ≥50 years old, 1224 were analyzed for their drusen and genotyped using illumina HumanOmniv3.5m or HumanHap610K. After imputation based on 1000 Genomes Project data and our standard quality control, 3,139,805 single nucleotide polymorphisms (SNPs) were included for the analysis. Large drusen was observed in 377 participants, while 837 did not have large drusen in their both eyes. We screened 19 AMD susceptibility regions reported by the AMD Gene Consortium with an adjustment for age and sex. P-value of <0.05 was observed in CFH, COL8A, CFI, IERD1R1, VEGFA, TGFB1, B3GALT1, RAD51B, LIPC, C3, APOE, TIMP3, and SLC16A8. In contrast to the strongest susceptibility effect of ARMS2/HTRA1 to AMD, SNPs in ARMS2/HTRA1 did not show significant associations to large drusen. Furthermore, SNPs in ADAMTS9, C2/CFB/SKIV2L, COL10A1, TNFRSF10A, and CETP did not show significant associations to large drusen. Stepwise analysis revealed 9 genes associated with large drusen; C3, CFH, TIMP3, LIPC, APOE, TGFB1R, B3GALT1, and IERF3DDR1. Our findings show some strong susceptibility genes for AMD do not show significant associations to large drusen. These findings suggest that roles of AMD susceptibility genes might be different in the early stage of AMD and in the late stage of AMD.
1067F

Present day limited resources demand DNA and phenotyping alternatives to the traditional prospective population-based epidemiologic collections. To accelerate genomic discovery with an emphasis on diverse populations, we as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study assessed all non-European American samples available in BioVU, the Vanderbilt University biorepository linked to de-identified medical records, for Metabochip genotyping. BioVU began DNA collection in the Davidson County area in 2007; as of 2011, a total of 15,863 samples were identified for Metabochip genotyping including African Americans (n=11,521), Hispanics (n=1,714), and Asians (n=1,122). Overall, more than half of the non-European descent patients are female (63.35%) and are young (mean age 37 years). The average number of clinical visits and billing (ICD-9) codes per patient was 81.8 (1 to 1,496) and 147.3 (1 to 3,817), respectively. The most frequent billing codes (collectively >50%) observed among African American adults >18 years reflected known population differences in prevalence and incidence of specific conditions such as hypertension, type 2 diabetes, and end stage renal disease. Despite the clinical origins of these data, comparisons with the population-based National Health and Nutrition Examination Surveys (NHANES) III and 1999-2002 (n=3,458 and 3,950 non-Hispanic blacks and Mexican Americans, respectively) suggest that the clinical genetics such as hypertension in African Americans (African American mean =27 kg/m² (EAGLE BioVU) and 27 kg/m² (NHANES)) and HDL-C (African American mean =50 mg/dL (EAGLE BioVU) and 51 mg/dL (NHANES)) exhibit little bias. Similar observations were made for BMI and HDL-C in EAGLE BioVU Hispanics versus NHANES Mexican Americans: 25.2 versus 27.1 kg/m² and 46 versus 47 mg/dL, respectively. Differences observed for mean LDL-C between BioVU (African Americans =99 mg/dL) and NHANES (non-Hispanic blacks =116 mg/dL) were attributed to well-documented lipid lowering medication usage in clinical practice. Collectively, these data suggest this clinical collection is essentially a random sample of the general population, making this resource more than comparable to traditional epidemiologic collections in the context of genetic association studies.

1069T
Mitochondrial genetic effects on proliferative diabetic retinopathy. D.C. Samuels1, M.A. Brantley, Jr2. 1) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Eye Institute, Vanderbilt University Medical Center, Nashville, TN, USA.

Diabetic retinopathy is the leading cause of vision loss in working-age adults in the US. The most severe form of this disease is Proliferative Diabetic Retinopathy (PDR), which is defined by aberrant blood vessel growth in the retina. The increased glucose level in diabetes causes increased flow of metabolites into the glycolytic pathway, while reactive oxygen species produced by mitochondria modulate the flow of metabolites out of the glycolytic pathway. The combination of these two effects, acting at opposite ends of glycolysis, determines the level of intermediate glycolytic metabolites and the rate at which these metabolites are diverted into alternative cytoxic pathways, such as the polyol, hexosamine, protein kinase C, and the advanced glycation end products (AGE) pathways. Increased activation of these pathways in diabetes has been suggested to cause cellular damage leading to diabetic retinopathy and other diabetic complications. We hypothesized that mitochondrial genetic variation alters reactive oxygen species production, thus modulating the flow of metabolites through glycolysis and affecting the development and progression of diabetic retinopathy. Using BioVU, we de-identified the version of the Vanderbilt Electronic Medical Record, we found that the common European mitochondrial haplogroup H was significantly overrepresented in the PDR cases compared to the Non-Proliferative Diabetic Retinopathy (NPDR) controls (p=0.038, OR=3.4 [95% CI 1.1-10.7], 21 cases & 54 controls). This result was replicated in a clinical cohort (p = 0.0024, OR = 3.8 [1.6-8.8], 58 cases & 44 controls). In addition, the common haplogroup U(k) was significantly protective against PDR in the combined BioVU and Clinical cohort group (p=0.031, OR = 0.43 [0.20-0.90]), though the association was not significant in the individual groups.

1070F
The association of 9p21.3 with acute myocardial infarction in managed care populations is independent of statin therapy. P. Erlich1, D. Carey2, S. Steinhubel3, C. McCarthy4, D. Cross4,5,1. 1) Center for Health Research, Geisinger Health System, Danville, PA; 2) Weis Center for Research, Geisinger Health System, Danville, PA; 3) Division of Research, EssentiA Institutes of Rural Health, Duluth, MN; 4) Marshfield Clinic Research Foundation, Marshfield, Wisconsin; 5) Department of Forensic and Investigative Genetics, School of Biomedical Sciences, University of North Texas Health Science Center, Fort Worth, Texas.

Introduction: A locus on 9p21.3 is associated with acute myocardial infarction (AMI) in multiple GWAS and replication studies. Current evidence suggests that 9p21.3 represents a previously-unknown pathway in AMI etiology, which is independent of traditional cardiovascular risk factors. If this hypothesis is true, it follows that the effect of 9p21.3 should not be influenced by the use of statins and that it should add a significant increment of performance to risk factor-based predictive models. Methods: This nested case-control study used electronic health records (EHR) linked to population-based biobanking initiatives of two integrated healthcare delivery systems in the US - Geisinger Health System in PA and Marshfield Clinic in WI. A cohort of 18,329 individuals (57% female; 108,400 person-years) aged 40 years and older was constructed. EHR were scanned to ascertain incident AMI status and covariates. Subjects with prevalent cardiovascular disease at baseline were excluded. A nested sample of 721 incident AMI cases and 722 matched controls was genotyped for a SNP (rs2383206) located in an intron toward the 3′ end of the CDKN2B-AS1 transcript. Bivariate and multivariate logistic regression was used to test association of rs2383206 with incident AMI and examine hypotheses related to effect independence. Receiver operating characteristic-area under the curve (ROC-AUC) was used to discern the net reclassification improvement of rs2383206 within a model with traditional risk factors. Results: The crude AMI incidence in the cohort was 10.8 and 5.0 per 1,000 person-years among males and females, respectively. In the nested sample, rs2383206 was associated with AMI in bivariate analysis (OR per A allele=0.71; 95%CI 0.62, 0.83; p<0.0001) and, with no notable difference in effect magnitude, in a multivariate analysis controlling for traditional risk factors and statin therapy (OR per A allele=0.72; 95%CI 0.61, 0.85; p<0.0001), suggesting an independent effect. The ROC-AUC of the parsimonious predictive model excluding rs2383206 was 70% (95%CI 67%, 73%), and increased by 1.2% upon addition of rs2383206. Conclusion: A SNP in 9p21.3 is associated with incident AMI in managed care populations independently of traditional risk factors and statin treatment. When added to a risk factor model, the SNP enhanced predictive performance by a small, statistically significant increment.
Clinical and genomic data, within a multidisciplinary collaboration to facilitate materials transfer contracts, the HDRC is making steady progress toward nation of a large number of sites and agreement among IRB approvals and suboptimal post-operative outcomes. A lack of standardized pathology and genetic findings into prognostic markers. How-ever, these results have been obtained from referrals of exceptional, not consensus for a given study requires a collaborative multidisciplinary team. Genetic epidemiologists and context experts define criteria for electronic phenotyping. Database queries are performed over the entire collection by software, and database developers to define phenotype groups. Laboratory values and longitudinal data are processed by statisticians to assess distribution assumptions and outliers. The final result is a set of samples with well-defined phenotypes for subsequent study. To date, this process has been completed for body mass index, type 2 diabetes and associated traits and conditions, cardiovascular disease and related traits, reproductive traits (ages at menarche and menopause), ocular diseases, peripheral artery disease, and kidney disease and associated labs. While new advances improve the ability to manipulate large datasets, these processes are computationally and labor intensive necessitating the creation of special workflows to transform clinical data into usable phenotypes for genomic discovery.

Hirschsprung Disease Research Collaborative (HDRC): A multidisciplinary partnership to advance Hirschsprung disease research. C. Berrios1, F. Abdullah1, P.K. Frykman2, R. Kapur1, J.C. Langer1, A. Chakravarti1, 1) Johns Hopkins University, Baltimore, MD; 2) Cedars-Sinai Medical Center, Los Angeles, CA; 3) Seattle Children’s Hospital, Seattle, WA; 4) Hospital for Sick Children, Toronto, Ontario, Canada.

Hirschsprung disease (HSCR) or congenital aganglionosis, a functional intestinal obstruction affecting 1 in 5,000 neonates, is a multifactorial neurodevelopmental disorder with >15 susceptibility genes identified and numerous more becoming evident through exome sequencing. Importantly, rare coding variants, common polymorphism and chromosomal anomalies, involving both coding and non-coding sequences, are all implicated in HSCR. Demonstrated correlation of variant types with clinical presentation offers promise for translation of genomic findings into prognostic markers. However, these results have been obtained from referrals of exceptional, non-random, cases. HSCR is treated surgically, and has a 30% incidence of suboptimal post-operative outcomes. A lack of standardized pathology and surgical approaches has prevented a full understanding of the variability in outcome. The HDRC process was computationally and labor intensive necessitating the creation of special workflows to transform clinical data into usable phenotypes for genomic discovery.


Variants in APOL1 are strongly associated with non-diabetic end-stage renal disease (ESRD) in African Americans (AA), however, the mechanism by which these variants affect renal function is unknown. APOL1 encodes basic leucine zipper (bZIP) transcription factors that regulate density lipoprotein (apo B-100) cholesterol. Given that patients with chronic kidney disease are prone to developing dyslipidemia, we investigated whether APOL1 variants were also associated with lipid levels. Additionally, it was of interest to obtain genetic risk estimates independent of hospital-based confounders. The HDRC is a repository of genetic data collected from patients and anonymously linked to their electronic medical records (EMRs). Three coding variants in APOL1 previously shown to be associated with ESRD in AA (Genovese et al. 2010) were genotyped in 13,327 Biobank participants (3706 AA, 4465 European Americans (EA), and 5356 Hispanic/Latinos (HL) by self-reported ancestry). MDRD estimated glomerular filtration rate (eGFR) and lipid levels were obtained from the patients’ medical records. Associations were analyzed using regression analysis with hospital-based model and adjustment for global proportion European, African, or Amerindian ancestry as an independent variable, using Illumina 770K chip data and the ADMIXTURE algorithm. The frequency of homozygous risk alleles at any of the three variants was 7.6% in AA, 0.7% in HLs, and 0% in EAs. Among non-diabetic patients, harboring two APOL1 risk alleles, AA were more likely to be hypertensive in both AA (OR 5.5, 95% CI 2.4-12.5, P=5x10-304) and HL (OR 21.9, 95% CI 2.3-207.2, P=0.007). Associations in AA, but not in HL, remained significant with adjustment for age, sex, blood pressure, BMI, and smoking status. The results were consistent with HLs and AA with hypertension.

Racial/ethnic differences for commonly measured clinical variables, such as cholesterol and blood pressure (BP), are well documented. Although the causes of these observed differences are unclear, it has been postulated that genetic factors may play a role. To test this hypothesis, we performed a genome-wide association study of triglyceride levels, non-HDL cholesterol, or total cholesterol in patients who were genotyped in 13,527 Biobank participants (3706 AA, 4465 European Americans, 6356 Hispanic/Latinos, and 2010) were genotyped on the Illumina 770K chip data and the ADMIXTURE algorithm. The frequency of APOL1 genotypes in 13,327 Biobank participants (3706 AA, 4465 European Americans (EA), and 5356 Hispanic/Latinos (HL) by self-reported ancestry).


Posters: Complex Traits and Polygenic Disorders

1071W Computational resources required to transform bedside data to base-pair research. W. Bush, J. Boston, E.Farber-Eger, R. Goodloe, D. Crawford, C. Berrios1, F. Abdullah1, P.K. Frykman2, R. Kapur1, J.C. Langer1, A. Chakravarti1, 1) Johns Hopkins University, Baltimore, MD; 2) Cedars-Sinai Medical Center, Los Angeles, CA; 3) Seattle Children’s Hospital, Seattle, WA; 4) Hospital for Sick Children, Toronto, Ontario, Canada.

1073F Risk assessment of APOL1 genetic variants in AA kidney disease. F. Abdullah1, J. Boston, E. Farber-Eger1, W. S. Bush1, 2, D. C. Crawford1, 3, 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 3) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN.

1074W Towards a pheno-wide catalog of human clinical traits impacted by genetic ancestry. L. Dumitrescu1, R. Goodloe1, J. Boston1, E. Farber-Eger1, W. S. Bush1, 2, D. C. Crawford1, 3, 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 3) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN.

Racial/ethnic differences for commonly measured clinical variables, such as cholesterol and blood pressure (BP), are well documented. Although the causes of these observed differences are unclear, it has been postulated that genetic factors may play a role. To test this hypothesis, we performed a genome-wide association study of triglyceride levels, non-HDL cholesterol, or total cholesterol in patients who were genotyped in 13,327 Biobank participants (3706 AA, 4465 European Americans, 6356 Hispanic/Latinos, and 2010) were genotyped on the Illumina 770K chip data and the ADMIXTURE algorithm. The frequency of APOL1 genotypes in 13,327 Biobank participants (3706 AA, 4465 European Americans (EA), and 5356 Hispanic/Latinos (HL) by self-reported ancestry). MDRD estimated glomerular filtration rate (eGFR) and lipid levels were obtained from the patients’ medical records. Associations were analyzed using regression analysis with hospital-based model and adjustment for global proportion European, African, or Amerindian ancestry as an independent variable, using Illumina 770K chip data and the ADMIXTURE algorithm. The frequency of homozygous risk alleles at any of the three variants was 7.6% in AA, 0.7% in HLs, and 0% in EAs. Among non-diabetic patients, harboring two APOL1 risk alleles, AA were more likely to be hypertensive in both AA (OR 5.5, 95% CI 2.4-12.5, P=5x10-304) and HL (OR 21.9, 95% CI 2.3-207.2, P=0.007). Associations in AA, but not in HL, remained significant with adjustment for age, sex, blood pressure, BMI, and smoking status. The results were consistent with HLs and AA with hypertension.

1075T Practical implementation of polygenic risk scores in personalized risk assessment for common cardio-metabolic traits. K. Fischer1, K. Läll2, J. R. Kőrös1, T. Haller1, T. Leitsalu2, T. Eisko2, A. Metspalu1, 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Institute of Mathematical Statistics, University of Tartu, Estonia.

Large-scale genome-wide association studies (GWAS) identify an increasing number of genetic variants associated with common complex diseases. Such markers (SNPs) are used to form polygenic risk scores. We will discuss the essential steps needed to translate GWAS results to personalized risk estimates. The process is illustrated using the data of Estonian population-based biobank and polygenic risk scores for Type 2 Diabetes (T2D), Coronary Artery Disease (CAD) and primary hypertension (HTN). The following questions will be addressed: 1)Validity of the genetic risk score in the target population. We will propose some graphical tools that address the question on consistency of published effect estimates (based on meta-analysis of a large number of cohorts) and estimates within the cohort of interest. That is done for each of the individual SNPs, as well as summary risk scores. 2)Conventional (non-genetic) risk factors and adjustment. Typically the GWAS results are obtained using minimal adjustment for other covariates. In practice, we are mainly interested in the added predictive value of a genetic risk score conditional on conventional risk factors. We will study the differences of conditional and unconditional effects, as well as possible heterogeneous across subsets with different baseline risk levels. 3)Number of markers to be included in the risk score. Increasing the number of SNPs does not necessarily lead to improved predictive accuracy in a particular cohort and there is an optimal number of SNPs producing the best results. A graphical tool is proposed to aid marker selection. 4)Population risk estimates from available genotyped samples. Often the genotyped (subsets of) biobank cohorts include a larger proportion of prevalent cases than the underlying population. We will show how the risk estimates can be corrected using external data on population prevalence. 5)Communication of the risk estimates in practice. We will propose some useful graphs that help to interpret the actual meaning of the risk estimate. For the Estonian Biobank cohort we show that the risk score for T2D has a very good discriminatory ability between high-risk and low-risk individuals, especially in those who are overweight, but not extremely obese (BMI 25-35). We will also show that the risk score for CAD should be mainly used to identify individuals at the highest percentiles of risk, whereas the risk score for HTN may be useful in some well-defined subsets of the cohort.


Growing numbers of biobanks hold large collections of clinically interesting samples. Common trends have emerged in needs for genotyping these large cohorts for biomedical studies, primarily the desire to simultaneously assay rare and common causative variants and large panels of common markers for genome wide association studies (GWAS). However, the detailed needs for each biobank or study depend on the ancestry of the cohort as well as the specific study aims. We present a system for optimizing an Affymetrix® Axiom® genotyping microarray for a specific study or biobank, including: (1) a computationally tractable method for selecting space-efficient probes to maximize genetic coverage in the population and allele frequencies of interest for GWAS; and (2) modular sets of markers relevant to specific biological and biomedical questions, which can be further customized for the specific phenotypes and populations of interest. We demonstrate the selection of only 246k markers, with minimal space requirements on the array, to provide 87% imputation-based coverage of the CEU population across 6.7M markers with MAF ≥ 5% in the 1000 Genomes Project (2012) genotype data. We also describe the selection of smaller sets of markers (‘booster panels’) to augment the coverage provided by this core 246k set in non-target populations, such as African and Asian populations, or to increase coverage of lower frequency markers. E.g., the addition of only 50k markers can be used to significantly raise imputation coverage of common Yoruba variants (MAF ≥ 5%), or increase coverage at lower minor allele frequencies in the primary target population, or raise coverage in any particular region of interest. Other sets of markers can be added to suit specific needs such as ADME markers or genetic markers known GWAS hits from previous studies, coding, non-synonymous SNPs and indels or markers with predicted loss-of-function impact, etc. This system for marker selection can improve the efficiency of genotyping in a biobank setting, potentially allowing more and more effective genetic studies in very large cohorts.

1077W ChiP-seq in alcoholic steatohepatitis and normal liver tissue identifies candidate disease mechanisms suggesting progression to cancer. C. Wadelius1, MB. Bysani1, O. Wallerman1, 2, S. Bomels4, K. Zatloukal1, J. Komorowski1, 2, 3, 4, 5. 1) Dept Immun, Gen & Pathol, Uppsala University, Uppsala, Sweden; 2) Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden; 3) Institute of Pathology, Medical University of Graz, Austria; 4) Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warszawa, Poland; 5) Current address: Department of Medical Biochemistry and Microbiology, BMC, Uppsala.

Background: Excessive consumption of alcohol may induce accumulation of fat in the liver concurrent with inflammation in a disease called steatohepatitis, which may progress to liver cirrhosis and hepatocellular carcinoma. Its molecular pathogenesis is largely made up of a few defined pathways. Histone modifications play a key role in transcriptional regulations as marks for silencing and activation of gene expression and for location of functional elements. Many transcription factors (TFs) are crucial for the control of gene activity and abnormality in their function may lead to disease. Methods: We performed ChiP-seq of histone modifications associated with active promoters and enhancers and a candidate transcription factor, in liver tissue from patients with steatohepatitis and normal liver and correlated results to mRNA expression and genotypes. Results: We performed ChiP-seq of the histone modifications H3K4me1, H3K4me3 and H3K27ac and the TF upstream regulatory site from 1077W patients in normal liver and from alcoholic steatohepatitis patients. We found many peaks that are differentially enriched between disease and normal tissue, and qRT-PCR results indicated that the expression of the tested genes strongly correlated with ChiP-seq of histone modifications but is independent of ≤3151 enrichment. By gene ontology analysis of differentially modified genes we found some genes which had previously been implicated in the etiology of steatohepatitis and others that are candidates to contribute to the disease. Importantly, the genes associated to the strongest histone peaks in the patient were over-represented in cancer specific pathways suggesting that the tissue was on a path to develop to cancer, a common complication to the disease. We also found novel SNPs and GWAS catalogue SNPs that are candidates to be functional and therefore needs further study. In summary we find that disease specific biomarkers and functional markers in tissue samples provides insights into disease mechanisms.

1078T National Biological Sample And Data Repository For WHO Group 1 Pulmonary Arterial Hypertension. M. Pascullo1, A. Reponen1, K. Lutz1, C. Winslow1, A. Walsworth1, J. Harley1, M. Barnes2, L. Martin1, K. Marsolo1, W.C. Nicholls1. 1) Division of Human Genetics, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH; 2) Division of Rheumatology, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH; 3) Division of Biomedical Informatics, Cincinnati Children’s Hospital, Cincinnati, OH.

We have established the National Biological Sample and Data Repository for PAH with funding from the National Heart, Lung, & Blood Institute of the National Institutes of Health. Biological samples, clinical data, and genotyping data are being collected on approximately 3,000 WHO Group 1 PAH patients from 3/2012 to 2/2017 to create this bio repository. Twenty four pulmonary hypertension centers across the United States have been enlisted to enroll patients, collect peripheral blood samples, and enter clinical data into a web-based eCRF. Many of the patients recruited for the bio repository are already enrolled in a multicenter, observational, U.S.-based Registry of PAH. Efforts are also being made to enroll treatment naïve patients to obtain both clinical and biospecimen samples. Treatment samples from serum and plasma are being isolated and banked using patient blood samples. Additionally, both DNA and RNA are being isolated from a portion of the obtained lymphocytes. Immortalized lymphocyte cell lines are also being established for each patient. Genetic data are being generated for each patient including genotypes for genome wide SNPs (5M) and coding sequence and/or MLPA data for BMPR2, ALK1, ENG, CAV1 and SMAD9. Additional genes can be added for screening as identified. All biological samples, clinical data, as well as the SNP genotype and sequencing/MLPA data for the patients will be made available to the research community using an online repository. The collected samples and data will become available for request after the first 500 patients have been enrolled and the data are completed. The National Biological Sample and Data Repository for PAH represents an unparalleled collaboration betweenADMEHDy modifiedcenters in the United States to enable the collection of the largest cohort of PAH patients’ biological samples in the United States. To date, 600 patients have been enrolled and sampled. This endeavor will provide the PAH research community an opportunity for here to pursue unprecedented hypothesis-driven studies. Funded by NHLBI HL105333.
1079F  Exome sequencing and genome-wide copy number variant mapping reveal novel associations with sensorineural hereditary hearing loss.  
R.R. Haraksingh1, F. Janehian Kenari2, J. Rodrigue-Paris2, J. Gelernter4, K. Nadeau1, J. Ogahali5, I. Schrijver5, M. Snyder2. 1) Psychiatry, Stanford University, Palo Alto, CA; 2) Genetics, Stanford University, Palo Alto, CA; 3) Department of Pathology, Stanford University School of Medicine, Stanford, California, USA; 4) Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; 5) Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, USA; 6) Department of Otolaryngology - Head and Neck Surgery, Stanford University School of Medicine, Stanford, CA 94305, USA.  
The genetic basis of hearing loss is not fully understood in terms of the diversity of loci and types of mutations that are responsible. We used multiple approaches, including exome sequencing of families and of probands with hearing loss, as well as copy number variation mapping in a case-control cohort, to identify loci associated with non-syndromic sensorineural hearing loss. Analysis of three distinct families revealed a novel gene candidate, MYH7B, associated with hearing loss in one family, and candidate loci in two other families. MYH7B encodes a Type II myosin, extending the role of cytoskeletal proteins in hearing. High-resolution genome-wide copy number variation analysis of 151 cases and 157 controls further revealed deletions in known hearing genes (e.g. GJB6, OTOA, and STRC, encoding connexin 30, otoancorin, and stereocilin, respectively), indicating that CNVs may be responsible for hearing loss. In addition, we found a novel association with hearing loss of a deletion on chromosome 16 containing the gene PDXDC1 (OR = 3.85, p = 1.45 × 10-7). Overall, our results indicate that a large number of loci, some of which are novel, and distinct types of mutations not typically tested for, may contribute to the etiology of hearing loss in humans.

1080W  Mutations in the BMP genetic network in patients with congenital GnRH deficiency.  
D. Cassatella1, J. Liang4, A. Dwyer5, G. Sykiotis1, H. Miraoui2, C. Xu1, S. Santini1, V.A. Hughes4, X.Z. Liu4, P.M. Bouloux2, M. Lang-Muntano6, R. Quinton4, J.G. Zhang1, B.J. Stevenson, Y. Sids1, N. Pitteloud1. 1) Service of Endocrinology, Diabetology and Metabolism - Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 2) VitalIT - Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) BGI-Shenzhen, Shenzhen, China; 4) Institute of Genetic Medicine, Newcastle University, Newcastle-upon Tyne, UK; 5) Centre for Neuroendocrinology, Royal Free and University College School of Medicine, London, UK; 6) Department of Endocrinology and Diabetology, University Children’s Hospital, Zurich, Switzerland.  
The hypothalamic secretion of gonadotropin-releasing hormone (GnRH) is well-known as the "pilot light" of reproduction in all mammals. GnRH neurons arise in the olfactory placode during embryonic development and migrate to the preoptic area of the hypothalamus. Mutations in genes involved in GnRH neuron ontology, survival and/or function result in GnRH deficiency (congenital hypogonadotropic hypogonadism, CHH), characterized by absent puberty and infertility. The mechanisms controlling GnRH neuron survival remain poorly understood. In this study, we analyzed variants in 17 BMP pathway genes including ligands (BMP2 and BMP4), receptors (BMPR1B and BMPR2), inhibitors (NOG and FST), and selected BMP4 as the best candidate to antagonizes Fgf8 signaling in various tissues, is expressed in the olfactory placode at E10.5. Heterozygous BMP4 mutations have been identified in 36 CHH patients, we employed a multi-step bioinformatics strategy incorporating network databases, and in silico functional predictions of variants. We found variants in 17 BMP pathway genes including ligands (BMP2 and BMP4), receptors (BMPR1B and BMPR2), inhibitors (NOG and FST), and enhancers (ENG and FBN1), and selected BMP4 as the best candidate to further study. In vitro studies revealed detrimental effects of BMP4 mutants on cell survival and protein expression. To strengthen the evidence that BMP4 is a novel locus for CHH, we are currently examining GnRH neuronal development in Bmp4 deficient mice. Studies of mutations in other BMP pathway genes are ongoing. In conclusion, we have identified mutations in the BMP network among patients with CHH. Impaired BMP signaling may disrupt the cross-talk between FGFS and BMPs in the olfactory placode during a critical stage for GnRH neuron development.

1081T  Identifying rare, non-coding DNA variants in Systemic Lupus Erythematosus.  
S.J. White1, S. Cantilliens4, E.F. Morand2. 1) Monash Institute of Medical Research, Monash University, Australia; 2) Southern Clinical School, Monash University, Australia.  
Systemic lupus erythematosus (SLE) is a debilitating multisystem autoimmune disease. Patients with SLE suffer a marked loss of life expectancy, and severe morbidity, due to autoantibody-mediated inflammation of multiple organs. A classic hallmark of SLE is the presence of autoantibodies. This is a clear biological link between SLE and antibody-producing B-lymphocytes (B-cells), strengthened by the clinical efficacy of anti-B cell therapies in certain patients. Genetic studies have associated SLE susceptibility with a number of Single Nucleotide Polymorphisms (SNPs), many of which are found in or near genes with B-cell functions. However, the known SLE SNPs together only account for a small percentage of the genetic contribution to SLE, and as most SNPs are located in non-coding DNA, the mechanism of their effect is often unclear. Several SLE-associated SNPs are located in such DNA elements that control gene regulation, with sequence changes disrupting the binding of specific transcription factors. In 2012 the ENCODE consortium published papers outlining genome-wide studies of open chromatin in a range of cell types. These studies demonstrated that DNAse hypersensitive sites strongly correlate with known and predicted regulatory sequences. In addition, they are enriched for known and predicted transcription factor binding sites (TFBS). This provides a clear link between open sequence variants and functional activity, and rare variants in B-cell regulatory loci (defined by open chromatin in B-cells) would explain genetic susceptibility to SLE in a way not detectable by GWAS. We undertook a pilot study to screen three B-cell regulatory elements previously associated with SLE. We identified previously undescribed sequence variant in DNA from an SLE patient, predicted to have a significant effect on transcription factor (TF) binding. This finding validates our hypothesis that non-coding regulatory elements will contain private sequence variants affecting TF binding sites and main structure, and may explain part of the ‘missing heritability’ issue in the post-GWAS era. We are currently performing additional mutation screening of putative B-cell regulatory loci in SLE patients, along with functional validation using reporter constructs and changes in DNase-sensitivity.

1082F  MTHFS Mutation May Contribute to Cerebral Folate Deficiency Syndrome.  
H. Zhu1, Y. Lei1, R.H. Fennell1. 1) Dell Pediatric Research Institute, Department of Nutritional Sciences, UT Austin, Austin, TX; 2) Department of Chemistry and Biochemistry, College of Natural Sciences, UT Austin, Austin, Texas.  
Cerebral Folate Deficiency (CFD) syndrome is characterized by very low concentration of 5-methyltetrahydrofolate (5-MTHF) in the patient’s cerebral spinal fluid (CSF), while folate levels in plasma and red blood cells are within normal limits. Major clinical features of CFD syndrome include unrelenting irritability, and insomnia, decelerating head growth, neurodevelopmental delay, regression, hypotonia, ataxia, dyskinesias, spasticity, speech difficulties, severe feeding problems, and variable behavioral issues. A specific folate pathway gene, 5,10-methenyltetrahydrofolate synthetase (MTHFS), associated with cerebral folate deficiency syndrome include: 1) Developmental delay, regression, hypotonia, ataxia, dyskinesias, spasticity, speech difficulties, 2) Severe feeding problems, 3) Variable behavioral issues. We are currently performing additional mutation screening of putative B-cell regulatory loci in SLE patients, along with functional validation using reporter constructs and changes in DNase-sensitivity.
Pilot whole genome sequencing of germline DNA from 186 breast cancer cases. P. Kraft, J. Allen, C. Chen, B. Decker, J. Figueroa, S. Hart, S. Lindstrom, J. Long, M. Yeager, S. Shanock, F. Couch, D. Easton, C. Haiman, W. Zheng, D. Hunter. 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK; 3) Division of Cancer Epidemiology and Genetics, NCI, Rockville, MD; 4) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 5) Division of Epidemiology, Vanderbilt University School of Medicine, Nashville, TN; 6) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 7) Department of Preventive Medicine, USC Keck School of Medicine, Los Angeles, CA.

High coverage whole genome sequencing (WGS) has been proposed to identify rare germline mutations associated with complex traits in both coding and regulatory regions. However, WGS studies present many logistical and analytic challenges, including coordinating variant calling and cleaning across multiple sites, accruing sufficient sample size, and designing and interpreting appropriate statistical analyses. We illustrate some of these challenges using a pilot whole genome sequencing study of 143 European ancestry (EA), early-onset, family-history-positive breast cancer cases, 21 Asian cases and 25 African-American cases from six studies participating in the DRIVE consortium, an NCI-sponsored post-GWAS initiative. Samples from each study were sequenced separately using Illumina HiSeq to an average depth of 30x and called individually using CASAVA. Preliminary analyses identified a missense BRCA1 mutation present in three EA cases that was absent in 4,300 EA subjects in the Exome Sequencing Project (p=6.8x10^-17). A nonsense variant in another gene was present in 11 cases by diploid (p=5x10^-14), but this apparent association was driven by false negative calls in the ESP—the average read depth at this position in the ESP was 1. We describe and present results from a rare-variant burden that uses summary data from individual studies. We discuss future plans, including combined variant calling and analysis of non-coding variants.


Hodgkin lymphoma (HL) shows strong familial aggregation but no major susceptibility genes for HL have been identified to date. Studies based on exome sequencing are promising for identifying disease susceptibility rare genetic variants. The goal of this study was to identify high-penetrance variants in HL-prone families. Using Nimblegen v2.0 and v3.0 exome capture array and the Illumina HiSeq2000 sequencer, we exome sequenced 45 HL cases or obligate carriers from 11 HL-prone families with three or more affected relatives. Variants found in more than 1% of samples from other studies identically processed and sequenced in our laboratory. The Ingenuity Variant Analysis (IVA) software on these newly recalled data will be presented. Additional analyses of these data show that HSCR can be sub-divided into four types arising from the effects of single genes, recurrent deletion/duplication loci, multi-organ syndromes of ‘multifactorial’ etiology and ‘multifactorial’ isolated cases, with effects, on average, from ~5 deleterious variants.

Analyses of WES data in multiplex Syrian non- endemic oral clefts families. J. Bailey-Wilson, E. Holzinger, M. Parker, S. Szymczak, Q. Li, C. Cropp, M. Nöthen, J. Hetmanski, H. Ling, E. Pugh, P. Duggal, M. Tsuchiya, I. Horvath, M. Marazita, M. Mangold, T. Beatty. 1) Inherited Disease Res Branch, NIH/NHGRI, Baltimore, MD, USA; 2) Department of Epidemiology, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA; 3) Institute of Human Genetics, University of Bonn, Bonn, Germany; 4) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD, USA; 5) Department of Biostatistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA; 6) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 7) Ibn Al-Nafees Hospital, Damascus, Syrian Arab Republic.

Oral clefts (cleft lip, cleft palate and cleft lip & palate) are common birth defects with a complex and heterogeneous etiology. Some genes and chromosomal regions have been associated with risk in genome wide association and linkage studies. This whole exome sequencing (WES) study used 22 affected 2nd degree or more distant relatives drawn from 10 multiplex inbred families (2 families with 3 relatives and 8 families with 2 relatives) initially ascertained in the Syrian Arab Republic for linkage studies. WES was done by the Center for Inherited Disease Research using the Agilent SureSelect v.4 capture reagents & Illumina HiSeq 2000 sequencers. Variants were called for all samples together within this project using Unified Genotyper (2.3-9). Variants were flagged by VQSR annotation using a Gaussian Mixture model for both SNVs and INDELs, Random Forests was used to estimate probability of high versus low quality calls. Results of analyses using the Ingenuity software on these newly recalled data will be presented. Additional sequencing studies of more families and more affected individuals in these families are ongoing to determine which genes segregate with oral clefts in these Syrian families.

The burden of coding, non-coding and chromosomal mutations in Hirschsprung disease. T. Turner, K. Nguyen, N. Krumm, S. Chattejee, A. Kapoor, Q. Jiang, A.Y. Ling, M.X. Sosa, N. Gupta, E.E. Eichler, S. Gabriel, C. Berrios, A. Chakravarti. 1) Center for Complex Disease Genetics, McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Genome Sequencing and Analysis Program, Broad Institute of MIT and Harvard, Cambridge, MA.

Hirschsprung Disease (HSCR) affects ~1/5000 live births and is characterized by the failure of gut innervation. This multi-factorial neurodevelopmental disorder has variable recurrence risks depending on the proband's sex, familiality and segment length of aganglionosis. 15 genes with rare, deleterious mutations in HSCR, most commonly at the receptor tyrosine kinase gene RET, are known. Although common functional variants in RET can enhance RET exist in ~90% of HSCR cases and all known rare coding functional variants in ~10% of patients, they cumulatively explain <10% of the risk variation. To understand the total burden of genomic variation, we performed exome capture (Agilent 44Mb) and sequencing on an Illumina platform in 304 patients representing the gamut of HSCR variability. We identify 90,818 high quality coding and splicing single nucleotide variants (SNVs) using GATK, and 146 rare (<1% in ESP) copy number variants (CNVs) using CoNIFER. Our major analyses have focused on 189 unrelated, European ancestry, non-admixed individuals and >361 analogous controls from dbGAP, ESP and 1000 Genomes. We focused on genes with 3 or more putative highly deleterious coding variants (NMD nonsense, donor/acceptor intronic splice, conserved (phyloP>4) missense) in cases and identified 64 genes with deleterious alleles at significantly higher frequencies as compared to the EVS. Among these, three genes are known to underlie HSCR in humans (RET, EDNRB, SEMA3D), one has dosage effects in patients (NAV2), and one causes aganglionosis in the mouse (PAX3). The novel genes show significant enrichment of genes in neuronal and/or ubiquitination pathways (22 or 57%) adding these to known HSCR effectors in enteric neural crest cell differentiation, neuronal function and axonal guidance. We find 11 CNVs, at ~10-fold increased frequency, at loci previously identified with recurrent deletion/duplication syndromes, such as 16p11.2, 1q21.1 and 17p11.2, suggesting a broader role for these loci in neuronal disease. Further analysis of these data show that HSCR can be subdivided into four types arising from the effects of single genes, recurrent deletion/duplication loci, multi-organ syndromes of `multifactorial’ etiology and ‘multifactorial’ isolated cases, with effects, on average, from ~5 deleterious variants.
Multigenic inheritance as a cause of familial congenital diaphragmatic hernia. T.F. Beck1, P.M. Campeau1, J.T. Lu2,2, C. Gonzaga-Jauregui2, J.R. Lupski1,4, R.A. Gibbs2, B.H. Lee1,5, W. Reardon6, D.A. Scott1,1, Centers for Mendelian Genomics. 1) Molec & Human Gen, Baylor College Med, Houston, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of Structural and Computational Biology & Molecular Biophysics, Baylor College of Medicine, Houston, TX; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Howard Hughes Medical Institute, Houston, TX; 6) Our Lady’s Hospital for Sick Children, Crumlin, Dublin 12, Ireland; 7) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX.

Congenital diaphragmatic hernia (CDH) is a common life-threatening birth defect which can present either alone or as part of a genetic syndrome. Isolated CDH has a low recurrence risk (1 to 2%) with most cases being caused by a combination of genetic and environmental factors. We report a male patient with isolated CDH whose father was also diagnosed with CDH at age 8 and exhibited severe scoliosis, low intellectual achievement, and unusual windsept hands. Both father and son had normal G-banded chromosome analyses. In an effort to identify the cause of CDH in this family, genomic DNA samples from the patient and his father were subjected to whole-exome sequencing analysis. This analysis revealed a frameshift mutation (c.4970_4971insA→p.Cys1658Leufs*29) in the fibrilin 1 gene (FBN1) which is associated with Marfan syndrome—a connective tissue disorder with a range of associated phenotypes, including lens dislocation, aortic dilatation, and skeletal anomalies. Sanger sequencing confirmed this mutation in father and son, as well as an asymptomatic male sibling who also shares his father’s unusual windsept hands. An ophthalmological evaluation revealed iridodonesis in both father and proband confirming the diagnosis of Marfan syndrome. The father also exhibited phacodonesis. A cardiovascular evaluation has been scheduled. Although CDH has been reported in a handful of Marfan cases, it is unusual for CDH to be the presenting symptom and to reoccur in more than one family member. With this in mind, we looked for mutations in other CDH-related genes which might account for the high prevalence of CDH in this family. Putatively damaging mutations were identified in the extracellular matrix protein 1 (FREM1), desmin (DES), and hepatocyte growth factor receptor (HGFGR) genes in both the father and his affected son. These findings underscore the importance of considering the potential influence of deleterious changes in other genes as an explanation for rare and/or unusual phenotypes associated with common genetic syndromes.

Exome sequencing identified novel genetic mutations in the patients with congenital vertebral anomalies. Y. Nakamura1, S. Kikugawa2, S. Seki3, M. Takahata4, H. Terai5, Y. Akaoka5, M. Matsubara6, F. Fujikoa6, H. Inagaki7, H. Kurahashi7, T. Kobayashi7, H. Kato1, 1) Shinshu University School of Medicine, Matsumoto, Japan; 2) DNA chip institute, Kanagawa, Japan; 3) Toyama Medical University, Toyama, Japan; 4) Hokkaido Medical University, Sapporo, Japan; 5) Osaka City Medical University, Osaka, Japan; 6) Nagano Prefectual Children’s Hospital, Azumino, Japan; 7) Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan; 8) Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

Purpose: Congenital Vertebral Anomalies (cVA) affect 0.5-1/1000 children and show a significant high mortality. Causes of most cVA remain largely unknown. The purpose of this study is to identify genetic abnormalities in Japanese cVA patients. Methods: We focus on primary cases with cVA that affect more than 1 vertebral body with malformation. Cases with congenital scoliosis with rib malformation, Klippel Feil syndrome, or VACTERL association were included in the patient cohort. We first performed exome sequencing on 8 patients and their 9 unaffected family members using the Agilent SureSelect Human All Exon 50Mb Kit and the illumina HiSeq Sequencer. Reads were mapped to the hg19 reference genome with Burrows-Wheeler Aligner and variant calling was performed by The Genome Analysis Toolkit. We excluded mismatches that did not change amino-acid sequences and known SNPs whose allele frequencies were greater than 1% according to the dbSNP JPT and 1000 Genome project. Approximately 700 such mismatches were found in each patient. We then confirmed the data by Sanger sequencing. A summary of results: We have found de novo mutations in the DCLRE1B, AGBL5, PDE2A, OLFML1, and ASB16 genes. All of these mutations are non-synonymous, and there has been no report of any mutations based on dbSNP and 1000 genome databases. DCLRE1B is involved in DNA interstrand cross-link repair; the major role of AGBL5 is tubulin deglutamylation; PDE2A is a phosphodiesterase involved in many signal transduction pathways; the function of OLFL1 is unknown. Although the function of ASB16 is mostly unclear, based on the sequence similarities and its domain structure, ASB16 may be a component of an E3 ubiquitin-protein ligase complex. Interestingly, a possible de novo nonsense mutation was found at a different position of the ASB16 gene in another patient. We are currently performing exome sequencing analyses on additional 4 families. We are also planning to create genetically-modified zebrafish models to understand the in vivo functional roles of these genes. Conclusion: These findings suggest that cVA are genetically heterogenous. Our study identified several potentially disease-causing genes of CVA.
Asthma is a complex genetic disease caused by a combination of genetic and environmental risk factors. We sought to test classes of genetic variants largely missed by whole-genome association studies (GWAS) by performing whole-genome sequencing of 16 individuals from asthma-enriched and depleted families from an extended 13-generation Hutterite pedigree with reduced genetic heterogeneity due to a small founding gene pool and reduced environmental heterogeneity as a result of a communal lifestyle. We sequenced each individual to an average depth of 13-fold (Illumina HiSeq), generated a comprehensive catalog of variant genotyped, and tested for association with asthma in 1200 individuals from the same population. Specifically, we identified and validated 2881 copy number variants (CNVs), 19 gene-disruptive single nucleotide variants (SNVs), and 18 frameshifting insertions or deletions (indels). We genotyped 593 of the CNVs in 1199 individuals, including 164 individuals with asthma, 488 controls, and 547 with intermediate phenotypes, from the extended pedigree. We identified a nominally significant association (p=0.03; Odds ratio = 3.13) between a 6 kb deletion in an intron of NEDD4L and increased risk of asthma. We genotyped this deletion in 742 cases and 755 controls of Puerto Rican ancestry and observed this variant in two cases and no controls. Interestingly, the NEDD4L is an ubiquitin ligase expressed in bronchial epithelial cells, and conditional knockout of this gene in the lung in mice leads to severe inflammation and mucus accumulation. This gene resides under a linkage peak for asthma in the Hutterites making it a plausible candidate. Additionally, this deletion is not in strong linkage disequilibrium with any particular SNP genotyped by the HapMap or 1000 Genomes Projects and would have been missed by GWAS. Of the 37 gene-disruptive SNVs and indels, we observed a reduced environmental heterogeneity as a result of a communal lifestyle.

Specifically, we identified and validated 2881 copy number variants (CNVs), 19 gene-disruptive single nucleotide variants (SNVs), and 18 frameshifting insertions or deletions (indels). We genotyped 593 of the CNVs in 1199 individuals, including 164 individuals with asthma, 488 controls, and 547 with intermediate phenotypes, from the extended pedigree. We identified a nominally significant association (p=0.03; Odds ratio = 3.13) between a 6 kb deletion in an intron of NEDD4L and increased risk of asthma. We genotyped this deletion in 742 cases and 755 controls of Puerto Rican ancestry and observed this variant in two cases and no controls. Interestingly, the NEDD4L is an ubiquitin ligase expressed in bronchial epithelial cells, and conditional knockout of this gene in the lung in mice leads to severe inflammation and mucus accumulation. This gene resides under a linkage peak for asthma in the Hutterites making it a plausible candidate. Additionally, this deletion is not in strong linkage disequilibrium with any particular SNP genotyped by the HapMap or 1000 Genomes Projects and would have been missed by GWAS. Of the 37 gene-disruptive SNVs and indels, we observed a reduced environmental heterogeneity as a result of a communal lifestyle.

Specifically, we identified and validated 2881 copy number variants (CNVs), 19 gene-disruptive single nucleotide variants (SNVs), and 18 frameshifting insertions or deletions (indels). We genotyped 593 of the CNVs in 1199 individuals, including 164 individuals with asthma, 488 controls, and 547 with intermediate phenotypes, from the extended pedigree. We identified a nominally significant association (p=0.03; Odds ratio = 3.13) between a 6 kb deletion in an intron of NEDD4L and increased risk of asthma. We genotyped this deletion in 742 cases and 755 controls of Puerto Rican ancestry and observed this variant in two cases and no controls. Interestingly, the NEDD4L is an ubiquitin ligase expressed in bronchial epithelial cells, and conditional knockout of this gene in the lung in mice leads to severe inflammation and mucus accumulation. This gene resides under a linkage peak for asthma in the Hutterites making it a plausible candidate. Additionally, this deletion is not in strong linkage disequilibrium with any particular SNP genotyped by the HapMap or 1000 Genomes Projects and would have been missed by GWAS. Of the 37 gene-disruptive SNVs and indels, we observed a reduced environmental heterogeneity as a result of a communal lifestyle.

Specifically, we identified and validated 2881 copy number variants (CNVs), 19 gene-disruptive single nucleotide variants (SNVs), and 18 frameshifting insertions or deletions (indels). We genotyped 593 of the CNVs in 1199 individuals, including 164 individuals with asthma, 488 controls, and 547 with intermediate phenotypes, from the extended pedigree. We identified a nominally significant association (p=0.03; Odds ratio = 3.13) between a 6 kb deletion in an intron of NEDD4L and increased risk of asthma. We genotyped this deletion in 742 cases and 755 controls of Puerto Rican ancestry and observed this variant in two cases and no controls. Interestingly, the NEDD4L is an ubiquitin ligase expressed in bronchial epithelial cells, and conditional knockout of this gene in the lung in mice leads to severe inflammation and mucus accumulation. This gene resides under a linkage peak for asthma in the Hutterites making it a plausible candidate. Additionally, this deletion is not in strong linkage disequilibrium with any particular SNP genotyped by the HapMap or 1000 Genomes Projects and would have been missed by GWAS. Of the 37 gene-disruptive SNVs and indels, we observed a reduced environmental heterogeneity as a result of a communal lifestyle.

Specifically, we identified and validated 2881 copy number variants (CNVs), 19 gene-disruptive single nucleotide variants (SNVs), and 18 frameshifting insertions or deletions (indels). We genotyped 593 of the CNVs in 1199 individuals, including 164 individuals with asthma, 488 controls, and 547 with intermediate phenotypes, from the extended pedigree. We identified a nominally significant association (p=0.03; Odds ratio = 3.13) between a 6 kb deletion in an intron of NEDD4L and increased risk of asthma. We genotyped this deletion in 742 cases and 755 controls of Puerto Rican ancestry and observed this variant in two cases and no controls. Interestingly, the NEDD4L is an ubiquitin ligase expressed in bronchial epithelial cells, and conditional knockout of this gene in the lung in mice leads to severe inflammation and mucus accumulation. This gene resides under a linkage peak for asthma in the Hutterites making it a plausible candidate. Additionally, this deletion is not in strong linkage disequilibrium with any particular SNP genotyped by the HapMap or 1000 Genomes Projects and would have been missed by GWAS. Of the 37 gene-disruptive SNVs and indels, we observed a reduced environmental heterogeneity as a result of a communal lifestyle.

Specifically, we identified and validated 2881 copy number variants (CNVs), 19 gene-disruptive single nucleotide variants (SNVs), and 18 frameshifting insertions or deletions (indels). We genotyped 593 of the CNVs in 1199 individuals, including 164 individuals with asthma, 488 controls, and 547 with intermediate phenotypes, from the extended pedigree. We identified a nominally significant association (p=0.03; Odds ratio = 3.13) between a 6 kb deletion in an intron of NEDD4L and increased risk of asthma. We genotyped this deletion in 742 cases and 755 controls of Puerto Rican ancestry and observed this variant in two cases and no controls. Interestingly, the NEDD4L is an ubiquitin ligase expressed in bronchial epithelial cells, and conditional knockout of this gene in the lung in mice leads to severe inflammation and mucus accumulation. This gene resides under a linkage peak for asthma in the Hutterites making it a plausible candidate. Additionally, this deletion is not in strong linkage disequilibrium with any particular SNP genotyped by the HapMap or 1000 Genomes Projects and would have been missed by GWAS. Of the 37 gene-disruptive SNVs and indels, we observed a reduced environmental heterogeneity as a result of a communal lifestyle.
Whole exome sequencing to identify variants influencing both pre-diabetic traits and type 2 diabetes mellitus in Pima Indians. L.J. Baier, K. Huang, A. Nair, Y.L. Muller, M. del Rosario, S. Kobes, R.L. Hanson, W.C. Knowler, C. Bogardus. PECRB, NIDDK/NIH, Phoenix, AZ.

Genes associated with type 2 diabetes (T2D) have been previously identified, but the causative variant underlying the association and the physiologic mechanism by which the gene influences T2D risk are often unknown. The goal of this study is to identify coding variation that increases susceptibility to T2D via its effect on a pre-diabetic trait. Exome sequencing was done on 177 Pima Indians and 148,616 variants (11.1% novel) were detected that met several quality and statistical criteria. Of 79 known variants and 15,842 were non-synonymous and synonymous SNPs, respectively. Selected variants (N=360) were genotyped in 555 non-diabetic Pima Indians characterized as inpatients in our Clinical Research Center for \% body fat (PFAT), central obesity (waist/thigh), glucose disposal rates (M) during a hyperinsulinemic-euglycemic clamp, acute insulin response (AIR) to intravenous glucose, and 2-hour plasma glucose concentrations (2-hr glucose) during an OGTT. Variants were further assessed for association with BMI and T2D in up to 7,667 subjects who had participated in a longitudinal study of T2D in the Gila River Indian Community. Two SNPs (in the genes ASXL3 and CYB5A) were significantly associated with a pre-diabetic trait after correction for multiple testing (360 SNPs analyzed for 5 traits required a p<2.7×10\(^{-7}\)).

Whole exome sequencing of 4,000 samples identifies rare variants strongly associated with type 2 diabetes risk in Mexicans and Latinos. K. Estrada\(^1,2,3\) for the SIGMA T2D Consortium. 1) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 3) Harvard Medical School, Boston, MA, USA.

Type 2 diabetes (T2D) is more prevalent in Latino populations than in people of European descent. We investigated the hypothesis that this increased T2D risk in Latino populations may be driven by rare, and possibly population-restricted, protein-coding variants that have remained undetected by previous studies focusing on common genetic variation.

We performed the largest ever whole-exome sequencing project of Mexicans and Latinos. We sequenced 1,946 samples equally distributed between T2D and controls. Samples were selected based on high Native American ancestry from a pool of 9,225 well-characterized individuals who had been genotyped with Omni 2.5 and ExomeChip arrays. After stringent quality control, 3,792 samples were available for analysis and included ~1.2M variants.

Whole Genome Sequencing to Identify Variants that Influence Pre-diabetic Traits in American Indians. K. Huang, P. Piacquio, S. Kobes, R. Hanson, C. Bogardus, L. Baier. Phoenix Epidemiology and Clinical Research Branch, NIDDK, NIH, Phoenix, AZ.

To identify genetic variation that influences pre-diabetic traits and thereby increases risk for type 2 diabetes (T2D), we obtained whole genome sequence data on 235 Pima Indians (127 men, age: 24.8±5.4 years, BMI: 33.4±7.4 kg/m\(^2\)). These individuals had been metabolically characterized for the following traits when they were non-diabetic: BMI, \% body fat, central obesity (waist/thigh), insulin-stimulated glucose disposal rate during a hyperinsulinemic-euglycemic clamp, acute insulin response to 25 g intravenous glucose bolus, and 2-hour plasma glucose concentrations during an oral glucose tolerance test. A subset of these subjects had also been profiled for gene expression levels (69 had expression data from a subcutaneous adipose biopsy and 90 had expression data from a subcutaneous adipose biopsy). Sequencing was performed by Illumina (N=200) and Complete Genomics, Inc (N=35). -12 million variants were found, including ~10.2 million SNPs, 58 million INDELs and 255,002 substitutions. Association analysis between variants and pre-diabetic traits is ongoing, as is analysis between variants and gene expression data to identify cis-acting variants. Variants with the strongest associations, as well as variants predicted to be damaging that map near candidate genes, will be genotyped in a sample of 555 Pima Indians who also have phenotypic information on pre-diabetic traits. Association with T2D will be assessed by genotyping in up to 7,667 American Indians who are part of a longitudinal study and have data on T2D status and BMI.
1096T
Whole exome sequencing identifies PAX4 nonsynonymous variant as susceptibility loci for type 2 diabetes in Koreans. S.H. Kwak1, J.I. Kim2, K. Kim1, Y.M. Cho1, H.S. Jung1, Y.J. Park1, K.S. Park1. 1) Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea; 2) Department of Biochemistry and Molecular Biology Seoul National University College of Medicine, Seoul, Korea.

Type 2 diabetes is a common complex disorder with strong genetic predisposition. Although more than 65 common genetic variants of diabetes have been identified so far, their effect sizes are small and they explain only a limited part of the heritability. In this study we used whole exome sequencing to identify low frequency, functional variants of type 2 diabetes. This was a case-control analysis using 324 confirmed type 2 diabetes patients and 101 carefully selected normal glucose tolerant elderly subjects without family history of diabetes. We limited our diabetes subjects to those who had at least one first degree relative of diabetes in their family. Whole exome capture was prepared using Agilent SureSelect version 4 + UTR and sequencing was performed by Illumina HiSeq 2000. Sequence alignment was done using BWA and Picard software and variant identification was done using GATK software. EPACTS software was used for rare variant association testing. A nonsynonymous variant in PAX4 (rs2233580, R192H) was associated with risk of diabetes in near exome-wide significance of \( P=9.66E-05 \) with minor allele frequency (MAF) in diabetes subjects 0.122 and controls 0.030. We then compared diabetes subjects having the rs2233580T variant of PAX4 gene (N=77) and diabetes subjects not having this variant (N=247). Those with the variant had lower body mass index (23.3±2.5 vs. 24.0±2.5, \( P=0.037 \)) and waist circumference (83.0±7.3 vs. 84.7±6.7, \( P=0.047 \)) with nominal significance. Another variant located in intron of SORL3 (rs12680280) was significantly associated with risk of diabetes with MAF 0.017 in diabetes subjects compared to 0.099 in control subjects (\( P=2.15E-08 \)). However, these variants require further replication and we are currently genotyping these variant in another set of 1,387 diabetes cases and controls and 821 gestational diabetes women. In conclusion, we have identified low frequency nonsynonymous variant in PAX4 to be associated with risk of diabetes in Koreans.

1097F
Large-scale exome chip association analysis identifies rare and low-frequency coding variants associated with glycemic traits. A. Mahajan1, X. Si1, A.K. Manning1, M.A. Rivas1, N. Grunewald2, H.K. Im3, H.M. Highland4, A.E. Locke5, P. Fontanillas5, T.M. Teslovich6, J. Flannick7, F. Cuschafer8, K. Gaulton9, H.M. Kang10, A.P. Morris11, J.B. Meigs12, C.M. Lindgren13 for T2D-GENES and GO-T2D Consortia. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxford, United Kingdom; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 4) Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK; 5) Department of Health Studies, University of Chicago, Chicago, USA; 6) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 7) Massachusetts General Hospital, Boston, Massachusetts, USA.

The extent to which low-frequency (LF; minor allele frequency (MAF) 1-5%) and rare (MAF <1%) coding variants contribute to variability of glycemic traits has not been systematically evaluated to date. To discover novel coding variants associated with fasting plasma glucose (FG) and fasting insulin (FI), and examine whether LF and rare coding alleles could explain established genome-wide association (GWA) signals for these traits, we studied up to 33,553 and 31,149 non-diabetic individuals of European ancestry for association with FG and FI respectively. Samples were genotyped using the Illumina HumanExome Beadchip array, which collectively represents >80% coding variation with >0.5% MAF in European ancestry populations. We tested variants for association with FG and FI using a linear mixed model to account for relatedness. We also carried out gene-based tests using the sequence kernel association optimal (SKAT-o) test. We then combined summary statistics at up to 135,904 high-quality autosomal variants and 14,667 genes across studies by meta-analysis. We identified one rare non-synonymous variant associated with FI at chip-wide significance (\( P=3.7x10^{-7} \)) in a gene not mapping to an established locus: 

\[
\text{COND} T2D-GENES: c.822C>T (p.L274P), 0.2\% \text{MAF; } T1278I
\]

and G6PC2 (\( P=2.75x10^{-7} \), 0.1% MAF; E594V). Two novel non-synonymous variants (one rare and one LF) were also associated with FG, approaching chip-wide significance (\( P=6.0x10^{-7} \); 0.2% MAF, T1278I) and GLP1R (\( P=1.5x10^{-7} \); 0.5% MAF, A319T). Of these, GLP1R and peptide-1 receptor, and therefore transduces incretin signaling in pancreatic beta cells. We also identified multiple rare coding variants in G6PC2, that in aggregate were associated with FG (SKAT-o \( P=5.31x10^{-15} \)), which reside within an established GWA locus. Furthermore, conditional single variant analysis confirmed the presence of two independent association signals in this gene, one common (\( P_{\text{COND}}=2.28x10^{-15} \); 48.2% MAF; V219L) and one rare (\( P_{\text{COND}}=1.23x10^{-10} \); 0.8% MAF; H177Y), neither of which can explain the GWA signal at this locus (lead SNP rs560887). They are also independent of nearby promoter variants (\( r^2<0.2 \)) identified in a recent study to be potentially functional. In conclusion, these results provide evidence that LF and rare coding variants contribute to variability in glycemic traits, and support G6PC2 as the causal gene at the established FG GWA locus.
**Posters: Complex Traits and Polygenic Disorders**

1098

Exome Chip genotyping in 9,000 individuals (type 2 diabetes and controls) in Mexican and Latinos. J.M. Mercader1,2, H. Moreno2, A. Huerta3, M.J. Gomez2 for the SIGMA T2D Genetics Consortium. 1) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA., United States of America; 2) Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston, Massachusetts, United States of America; 3) Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran, Mexico City, Mexico.

Type 2 diabetes (T2D) has become a mounting health problem in Mexico, with an estimated nationwide 25% increment in prevalence over a 7 years period. To date, most genome-wide association studies (GWAS) to identify susceptibility genes for T2D have been carried through common variant genotyping arrays with very comprehensive coverage of European variability, but with lower coverage in other populations. Under the hypothesis that population-specific common and rare coding variants, in Mexicans and Latin-American may influence risk of T2D, we genotyped 4,210 T2D cases and 4,786 controls with the Exome Chip. The Exome Chip was designed to capture 235,933 protein-coding variants (mostly non-synonymous, splice-site and stop altering SNPs) that were selected based on ~12,000 sequenced exomes (including hundreds of Mexican American samples), as well as some common variants to control for population stratification and to tag already known GWAS hits. After quality control, based on several metrics and on genotype concordance with OMNI2.5M array in the same samples as well as a subset of 4,000 samples with exome-sequence, 161,676 variants in 8,621 individuals were tested for association using logistic regression adjusting for body mass index, age and ancestry via 10 principal component analysis. Among the top hits, we replicated the previously reported associations of SNPs in or near TCF7L2, and KCNQ1, with similar effect sizes and direction of effects shown in other populations. We also validated those variants identified last year in SLC16A1 (SIGMA T2D Genetics Consortium, submitted) as well as an association in a splice-site acceptor variant in the INS-IGF2 gene (OR=0.79; p=7.0x10^-14) found through imputation in the same dataset. This variant, not yet replicated, has only been reported in Latin Americans, with its highest frequency in Mexico (MAF=0.17), and East Asians (MAF=0.07) (Methodology). According to this study, a test statistic (P-value) of 0.05 is sufficient to claim evidence. The most compelling result emerging from 11,497 subjects from the CoLaus and LOLIPOP studies was an association between rs10305492 coding for GLP1R and fasting glucose (adjusted for BMI), the target of the GLP1 agonists, with a p value of 7x10^-5. It has been anticipated that mining of the low frequency spectrum of variation will identify meaningful associations that were not observed with common variation (>5% MAF). With a focus on type 2 diabetes and related traits including obesity, we attempted to replicate the previous findings from a large sequencing study of 202 drug-target genes in 4065 subjects phenotyped for cardiovascular and metabolic traits (Nelson 2012). We selected 30 variants to be assessed in up to 46,037 subjects of European origin (cohorts including CoLaus, LOLIPOP, EPIC-Norfolk, Norfolk Diabetes Study, Ely and Fenland). All variants identified in the initial study were also imputed into a subset of these studies where genome-wide data were available (up to 10k individuals). We performed meta-analyses to test for association between these 30 variants and diabetes, obesity, glucose and obesity-related traits. A p value of <3.6x10^-5 was considered significant (independent tests for 202 genes for a given trait). Imputed and genotyped results of interest included SNPs within the GLP1R gene associated with fasting glucose (p=4x10^-4), HRH1 gene variants associated with insulin (p=2.7x10^-5), IL1R1 (p=2e-4) variants associated with 2-hr glucose, CNR2 variants associated with HOMA in obese subjects (p=3e-4) and NTRKB variants associated with waist (4e-4). These results were not previously identified by genome-wide studies at genome-wide significant levels. Additional results from recently published studies are being prioritized for follow up. We conclude that Exome chip genotyping exomes (including hundreds of Mexican American samples) provides a powerful tool to study new genetic variants influencing diabetes and related traits.

1099

A low frequency coding variant (A316T) in the glucagon-like protein receptor 1 (GLP1R) is associated with fasting glucose levels. D. Waterworth1,2, R. Scott1,2, L. Lii1, C. Gillson1, J. Aponte1, L. Warren2, S. Chissone3, M. Ehm3, N. Wareham1, 1) Genetics, GlaxoSmithKline, King of Prussia, PA; 2) Department of Public Health, University of Cambridge, UK; 3) Statistical Genetics, GlaxoSmithKline, RTP, NC, 4) Genetics, GlaxoSmithKline, RTP, NC.

It has been anticipated that mining of the low frequency spectrum of variation will identify meaningful associations that were not observed with common variation (>5% MAF). With a focus on type 2 diabetes and related traits including obesity, we attempted to replicate the previous findings from a large sequencing study of 202 drug-target genes in 4065 subjects phenotyped for cardiovascular and metabolic traits (Nelson 2012). We selected 30 variants to be assessed in up to 46,037 subjects of European origin (cohorts including CoLaus, LOLIPOP, EPIC-Norfolk, Norfolk Diabetes Study, Ely and Fenland). All variants identified in the initial study were also imputed into a subset of these studies where genome-wide data were available (up to 10k individuals). We performed meta-analyses to test for association between these 30 variants and diabetes, obesity, glucose and obesity-related traits. A p value of <3.6x10^-5 was considered significant (independent tests for 202 genes for a given trait). Imputed and genotyped results of interest included SNPs within the GLP1R gene associated with fasting glucose (p=4x10^-4), HRH1 gene variants associated with insulin (p=2.7x10^-5), IL1R1 (p=2e-4) variants associated with 2-hr glucose, CNR2 variants associated with HOMA in obese subjects (p=3e-4) and NTRKB variants associated with waist (4e-4). These results were not previously identified by genome-wide studies at genome-wide significant levels. Additional results from recently published studies are being prioritized for follow up. We conclude that Exome chip genotyping exomes (including hundreds of Mexican American samples) provides a powerful tool to study new genetic variants influencing diabetes and related traits.

1100

Whole genome sequencing identifies novel low frequency variant associations in liver function traits. L. Quaye on behalf of the UK10K Consortium Cohorts Group. Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom.

1,754 individuals from the TwinsUK cohort were whole genome sequenced (WGS) to average 6x coverage using next-generation sequencing technology, as part of UK10K Consortium Cohorts Group. Variants from 1000 Genomes and those discovered through WGS conducted on 3,621 individuals within UK10K were imputed into individuals in the TwinsUK cohort who had been previously genotyped, thereby increasing the sample size for analysis. In a meta-analysis of directly sequenced and imputed variants, we evaluated variants for association with liver function traits (albumin, alkaline phosphatase, total bilirubin and gamma-glutamyl transpeptidase (GGT)) of individuals from the TwinsUK cohort. Association analysis was performed on quantile-normalised abundance of the liver function traits, with adjustments for age and body mass index. 22 previously reported genome-wide significant associations between total bilirubin and the UGT1A cluster on chromosome 2 were replicated (P<5e-8). In addition, an association was identified between levels of bilirubin and a variant of KCNB1 (20q13.2) with a minor allele frequency (MAF) of 0.026 (beta= 0.61, standard error (SE)=0.11, P=6.02e-08). This variant, also associated with height, is a missense mutation within the coding region, resulting in a proline to serine amino acid substitution. This association is being replicated with exome sequencing data and an independent cohort. A further genome-wide significant association was found between GGT and a rare variant not available in HapMap2 reference panels, near TEC14, on chromosome 17p13 (MAF=0.019; beta=-0.828492; SE=0.17, P=6.99E-7), which is not available on Hapmap2. A SKAT analysis of 26,367 gene regions using 42,799 non-overlapping windows with maximal 50 coding variants with MAF<0.05 showed a suggestive association (P=3.1E-5) between KCNB1 and bilirubin. In conclusion, we found between bilirubin and GGT and variants with MAF<0.05. These promising early results from the UK10K project highlight the value of WGS studies in identifying novel associations and indicate that low frequency variants are likely to contribute to the genetic variance of liver function traits.
1101W
Autozygosity Mapping in Pakistani Intellectual Disability Families. M. Rafiq1, K. Mittal1, L.A. Balouch1,2, A. Noor1,2, C. Windpassinger4, A. Mikhailov1, M. Aslam3, M. Ayaz4, A. Mi5, M. Ansari2, P. John2, M. Ayub2, J.B. Vincent1. 1) Molecular Neuropsychiatry, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Atta-ur-Rehman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad-Pakistan; 3) Dept. of Pathology & Laboratory Medicine, Hospital for Sick Children, Toronto, Canada; 4) Institute of Human Genetics, Medical University of Graz, Austria; 5) Lahore Institute of Research & Development, Lahore, Pakistan; 6) International Islamic University, Islamabad, Pakistan; 7) Dept. of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan; 8) Division of Developmental Disabilities, Dept. of Psychiatry, Queen’s University, Kingston, ON, Canada.

Autosomal recessive causes of intellectual disability (ARID) have, until very recently, been under-researched due to the high degree of genetic heterogeneity. However, now that genome-wide approaches can be applied to single multiplex consanguineous families, identification of genes harboring disease-causing mutations by autozygosity mapping is expanding rapidly. We have ascertained more than 165 multiplex, consanguineous ARID families from Pakistan. These families are selected for lack of obvious syndromic features. Our strategy includes genotyping family members on genome-wide single nucleotide polymorphism microarrays, looking for large regions of shared homozygosity (and haploidentity) between affected individuals (homozygosity-by-descent, or autozygosity). We also screen for potential disease-related CNVs- either as a shared homozygous genotype, or heterozygous as a potential cause of phenocopy. We firstly exclude any known ARID genes in HBD regions, then either select candidates from within the HBD region for mutation screening by Sanger sequencing, or we embark on whole exome sequencing to identify disease mutations. Our successes include a number of new genes for apparent non-syndromic ARID, such as MAN1B1, TRAPPC9, NSUN2, as well as new genes for syndromic forms or ARID, such as Joubert syndrome (CC2D2A and TCTN2), and many known ARID genes (TUSC3, TPO, VPS13B, PEX1, PSPH, PMM2). Here we describe the use of autozygosity mapping and whole exome sequencing to identify an additional 6 new genes for NS-ARID. As more and more genes for ID are defined, the path with which we build a picture of the biological pathways that, when perturbed, may lead to intellectual disability.

1102T
Identifying Genetic Variants Associated with Anorexia Nervosa via Exome Sequencing. S. Yu, E. Pruett, R. Cone, B. Li. Vanderbilt University, Nashville, TN.

Anorexia nervosa (AN) is a highly heritable psychiatric disease characterized by inability to maintain a minimal normal weight, persistent fear of gaining weight and preoccupation about body shape. It affects 0.5-1% of women and 0.3% of men in the population with females 10 times more likely to be affected than males. The estimated heritability of AN is 56-75%; however the genetic basis of AN is largely unknown. Results from previous candidate gene studies are often not replicated and genome wide association studies focusing on common variants (<5%) do not reveal convincing association signals. We hypothesize that low frequency and rare variants associated with AN etiology. To identify genes and pathways harboring such high-risk variants, we describe the use of autozygosity mapping and whole exome sequencing to identify disease mutations. Our successes include a number of new genes for apparent non-syndromic ARID, such as MAN1B1, TRAPPC9, NSUN2, as well as new genes for syndromic forms or ARID, such as Joubert syndrome (CC2D2A and TCTN2), and many known ARID genes (TUSC3, TPO, VPS13B, PEX1, PSPH, PMM2). Here we describe the use of autozygosity mapping and whole exome sequencing to identify an additional 6 new genes for NS-ARID. As more and more genes for ID are defined, the path with which we build a picture of the biological pathways that, when perturbed, may lead to intellectual disability.

1103F

Neural tube defects (NTDs) are among the most common of birth defects, occurring in approximately 1 in 1000 live births in the US. Caused by improper closure of the developing neural tube during development, NTDs encompass a broad range of phenotypes that are likely to have a complex etiology with both genetic and environmental factors. The most severe NTD phenotype is anencephaly, an open NTD that results in the partial or complete loss of the brain and skull, and is therefore incompatible with life. While most cases are sporadic, some families have multiple affected offspring and are more likely to harbor a highly penetrant genetic risk factor. We performed exome sequencing on one such family who had three consecutive anencephalic pregnancies and a fourth unaffected pregnancy. We have completed the sequencing and analysis of the parents and the first two affecteds in this multiplex family. While we observed that the second affected sample was contaminated with maternal tissue (likely taken at the time of collection), we were able to correct for this using parental genotypes and allele proportions at each site. Exonic regions were enriched using the Illumina TrueSeq Exome Enrichment kit, sequenced 2x100nt on an Illumina HiSeq2000 instrument, and aligned against NCBI build 37. Variants were called using the Genome Analysis Toolkit, and only sites with a minimum read depth of 3, quality score of 50, mapping quality of 45 and quality by depth of 7 were included in analysis. The initial set of 115,066 variants was further filtered by requiring concordant genotypes and the presence of at least one minor allele in both affecteds; 34,701 variants met these criteria. We identified 10 nonsynonymous variants in genes belonging to the Gene Ontology term cilium morphogenesis. All of these variants had deleterious effects as predicted by PolyPhen-2, SIFT, and/or PROVEAN, including a pair of compound heterozygous sites in the gene KIF24. These two variants had minor allele frequencies of 0.8% and 19.1% and were inherited from both the mother and father. Exome sequencing of the third affected and the unaffected sibling are currently underway to refine the putative causal genetic risk factors in this family. We are encouraged that the initial findings have implicated a novel candidate gene and that additional functional analyses will be necessary to fully delineate the role of these polymorphisms in the development of anencephaly.

1104W
Whole exome sequencing case-control using 1,000 severe obesity cases identifies putative new loci and replicates previously established loci. A. Hendricks on behalf of the UK10K Consortium: Obesity. Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The UK10K project (www.uk10k.org) has sequenced 10,000 individuals: 4,000 cohort participants using low depth (6x) whole-genome sequencing (WGS) and 6,000 disease cases using high depth (50x) whole-exome sequencing (WES) with the aim of providing a variant resource to the community and identifying new disease loci. Of the 6,000 WES cases, 1,000 are patients from the Severe Childhood Onset Obesity Project (SCOOP). SCOOP consists of UK white patients, with an age of onset below 10, and with a body mass index (BMI) Standard Deviation Score (SDS) > 3, a design that is enriched for rare, highly penetrant causes of obesity. Single-variant and gene-region based case-control analyses, using two different control sets, has confirmed previously known obesity signals as well as identified potentially new variants and gene-regions associated with severe childhood obesity. The single-variant case-control analyses were performed using low depth WGS UK10K population cohort, and a subset of the other WES UK10K disease cases that had been consented for use as controls. After thorough pre- and post- analysis quality control, the results do not show any apparent systematic bias and reproduce 27 promising single variant associations (p-value < 1e-5), which we are following-up in additional samples. Amongst these is a single variant association within the ADCY3 region (p=3.6e-06), a gene with an established association with BMI. Finally, we performed gene-region based case-control analysis using UKAT-O using the other WES UK10K disease cases that had been consented for use as controls. We find several potentially novel gene-regions associated with obesity and are currently performing targeted resequencing of nearly 3,000 independent SCOOP samples to replicate and add further evidence to these findings. The case-control analyses and replication efforts described here highlight the potential value of case-control using WES data to identify disease-associated loci and will help gain further insight into severe childhood obesity.
1105T

Rare variant association analysis reveals novel associations with lipids in genes within established loci via imputation up to the 1000 Genomes Project reference panel. R. Magi, M. Horikoshik, I. Surakkaik, S. Gustafssonik, S. Häggj, C. I. Ladenvallk, L. Marullo2,3, C. P. Nelson4,5, J. S. Ried2, J. Thorleifsson2, N. Tsimikovak, S. M. Willemesk, C. Willenborg1, T. Winkler2, C. M. Lindgren2,6, M. I. McCarthy2,3, S. Ripatti4,5, I. Prokopenko2,6, A. P. Morris2, ENGAGE consortium. 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Oxford Centre for Diabetes, Endocrinology and Metabolism; University of Oxford, Oxford, UK; 4) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland; 5) National Institute for Health and Welfare, Helsinki, Finland; 6) Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, Netherlands; 7) Netherlands Consortium for Healthy Ageing, Leiden, Netherlands; 8) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 9) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University and Lund University Diabetes Centre, CRG at Skåne University Hospital, Malmö, Sweden; 10) Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy; 11) Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; 12) National Institute for Health Research (NIHR) Leicester Biomedical Research Centre; 13) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 14) deCODE Genetics, Reykjavik, Iceland; 15) Department of Genetic Epidemiology, Erasmus MC, Rotterdam, The Netherlands; 16) Karolinska University Hospital, Stockholm, Sweden; 17) Public Health and Gender Studies, Regensburger University Medical Center, Regensburg, Germany; 18) Broad Institute, Boston, MA, USA; 19) Genomics of Common Disease, Imperial College London, London, UK.

Genome-wide association studies (GWAS) have been extremely successful in identifying novel loci contributing genetic effects to plasma concentrations of lipids. Two loci (at PHCAP and 7p21.1) have previously been implicated in hypertriglyceridemia. By imputation into existing GWAS data up to high-density reference panels as an alternative to costly re-sequencing experiments, a tool for discovering rare variant associations, and an imputation into existing GWAS data up to high-density reference panels as an alternative to costly re-sequencing experiments, we can identify variants associated with complex traits. Our results highlight the potential of imputing to high-density reference panels as a tool for discovering rare variant associations with complex traits, and an alternative to costly re-sequencing experiments.

1106F

Identification of common and rare variants associated with trunk fat mass using whole-genome sequencing in the UK10K project. I. Paternoster, UK10K Consortium Cohorts Group. MRC Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, United Kingdom.

As part of the UK10K Consortium, 1867 individuals from the Avon Longitudinal Study of Parents & Children (ALSPAC) were whole genome sequenced (WGS) to ~6.5x coverage. Varians discovered through WGS, along with those from 1000 Genomes were imputed into the full cohort with genome-wide SNP data, increasing the sample size to 5498. Trunk fat mass (a highly accurate measure of central obesity) was derived at age 9yrs using dual energy x-ray absorptiometry (DXA). To investigate common variation, we tested 8.8 million variants (MAF>1%) in a meta-analysis of several and imputed variants for association with trunk fat mass. We found three loci with p<5×10^-8. Two of these are in known BMI regions (ADCY3 & FTO). The other locus (7p21.1) has previously shown suggestive evidence for association with BMI 2-score change in Hispanic children (p=7×10^-6). We also found a modest association signal between this variant and BMI in our data (p=7×10^-6), but no association with waist hip ratio (WHR). The GIANT GWAS of BMI showed no association with this variant (p=0.847) which indicates that this signal might be specific to childhood obesity. 25 loci were suggestive associated (p<1×10^-4); 17 are novel for weight-related traits and 9 are variants of low frequency (MAF<5%). We are replicating these in the TwinsUK arm of the UK10K project (n=100 sequenced) and additional child cohorts. For rare variants (in WGS data only) we used SKAT to analyse 26,367 gene regions using 41,692 non-overlapping windows with maximal 50 coding variants with MAF<5%. We found three regions (IL17RC, PPP4R1 & RP11-115285) associated with trunk fat mass (p<1×10^-5). Further investigation of the signals in these regions show that two of these (IL17RC & RP11-115285) are driven by only one variant, whereas the signal for PPP4R1 appears to be driven by two variants. Two of these genes (IL17RC & PPP4R1) also showed suggestive evidence for association with BMI (p<0.01), but no association with WHR. We are following these up by genotyping in the whole of the ALSPAC cohort to allow validation and replication. We will also investigate enrichment of hits according to functional annotation. The UK10K WGS project has identified novel loci for trunk fat mass. Our findings suggest that both rare and common variants contribute to the variance of this trait.
1107W

Statistical dissection of genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1) using whole-genome sequencing data. R. Ruizc1, M. Almeida1, E. Dingalenko1, T.D. Dyer1, T.M. Teslovich2, G. Jun3, J.M. Peralta3, C. Fuchsberger2, A.R. Wood4, A.R. Manning4, T.M. Frayling2, P. Cingolani3, R. Sladek2, D.M. Lehman, J.W. Kent Jr.1, J.B. Harley5, M.A. Carless1, J.E. Curran5, M.P. Johnson, S.A. Cole1, L. Almasy4, E. Kraig5, G. Abecasis1, R. Yoken5, R. Duggirala1, C.T. Leach2, J. Blangero1, H.H.H. Göring1, T2D-GENES Consortium. 1) Dept of Genetics, Texas Biomed, San Antonio, TX; 2) University of Michigan, Ann Arbor, MI; 3) University of Exeter, Exeter, United Kingdom; 4) Broad Institute, Boston, MA; 5) McGill University, Montreal, Canada; 6) Montreal Diabetes Research Institute, Montreal, Canada; 7) University of Texas Health Science Center at San Antonio, San Antonio, TX; 8) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 9) Johns Hopkins University School of Medicine, Baltimore, MD.

Infection with Epstein-Barr virus (EBV) is highly prevalent among populations around the world and can cause infectious mononucleosis and more severe diseases including Hodgkin lymphoma, nasopharyngeal lymphoma, and post-transplant lymphoproliferative disorders. We measured IgG antibodies against EBV nuclear antigen 1 (EBNA-1) in extended Mexican American families from San Antonio, Texas. We previously performed association analysis in 1,956 individuals using 1M SNPs and reported significant association in the human leukocyte antigen (HLA) region on chromosome 6. There were at least two independent loci in this region (lowest p-value of 1.4×10^{-15}), with HLA class II genes HLA-DRB1 and HLA-DQB1 identified as the best candidates. This region was not associated with antibody titer levels against HLA-DRB1 identified as the best candidate in the human leukocyte antigen (HLA) region on chromosome 6. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some study participants as part of the T2D Consortium.

Posters: Complex Traits and Polygenic Disorders

1108T

Whole-exome sequencing in age-related macular degeneration. P. Whitehead1, W.K. Scott1, G. Wang2, W. Cade3, M.D. Courtenay1, S.G. Schwartz4, J.L. Kovach1, A. Agarwal1, J.L. Haines2,4, M.A. Pericak-Vance1, 7, A.R. Wood4, A.R. Manning4, T.M. Frayling2, P. Cingolani3, R. Sladek2, D.M. Lehman, J.W. Kent Jr.1, J.B. Harley5, M.A. Carless1, J.E. Curran5, M.P. Johnson, S.A. Cole1, L. Almasy4, E. Kraig5, G. Abecasis1, R. Yoken5, R. Duggirala1, C.T. Leach2, J. Blangero1, H.H.H. Göring1, T2D-GENES Consortium. 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 3) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 4) Ophthalmology and Visual Sciences, Vanderbilt University School of Medicine, Nashville, TN.

Genome-wide association meta analysis (GWAMA) has implicated common variants in 19 genes associated with AMD, including variants in FSHR, CD276, PLD5, KRT26, and KRT27. We have previously reported well captured by genome wide association studies and may represent some of the unexplained heritability in AMD risk. We used whole exome sequencing (WES) in phenotypically extreme individuals to identify RVs implicating novel AMD genes. The GWAS 19 loci were used to calculate a genetic risk score (summed number of risk alleles weighted by effect sizes) for each individual. We created an ‘extreme’ case/control sample with the following characteristics: 38 individuals with bilateral neovascular AMD and the lowest calculated genetic risk score, and youngest ages at examination; 37 unaffected controls with no drusen, the highest calculated genetic risk score, and oldest ages at examination. Sequecing Capture was by Agilent SureSelect. Alignment and base calling used the Illumina CASAVA 1.6 pipeline, aligned to hg19 using BWA. Single nucleotide variants (SNV) and insertion-deletion variants (indels) were called by GATK Unified Genotyper with VQSLAR alignment. Variants with VQSLOD<3 and variant genotype likelihoods <99 were excluded. All variants were annotated using SeattleSeq Annotation. Association of individual SNV with AMD was assessed by Fisher’s exact test (SAMD). P-values were corrected for multiple testing using the false discovery rate (FDR) method of Benjamini-Hochberg. 19 genes were used to define genotypic extremes were excluded from the gene-based analysis. No gene-based tests met Bonferroni-corrected significance (p<5×10^{-4}). When considering only rare variants, 19 gene based test p<5×10^{-4}. MOG is close to known AMD gene IER3, and others reported for AMD. 10 SNV in the known AMD loci were associated with the extreme phenotype. Initial results of this WES suggest additional genes with multiple rare SNVs may influence AMD.

1109F

Whole Genome Sequencing Association Study for Quantitative Ultrasound of the Calcaneus. S.G. Wilson1,2,3,4, on behalf of the UK10K cohorts. 1) Endocrinology & Diabetes, Sir Charles Gairdner Hosp, Nedlands, Australia; 2) Department of Twin Research, King's College London, London, UK; 3) Medicine & Pharmacology, University of Western Australia, Australia; 4) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK.

Aim: To identify genetic variants associated with Quantitative Ultrasound (QUS) of the Calcaneus which is used for evaluation of osteoporosis, we performed an association study using whole genome sequence (WGS) data from the UK10K Consortium.

Methods: 1327 individuals from TwinsUK cohort with QUS phenotypes (broadband ultrasound attenuation (BUA) and velocity of sound (VOS)) and WGS were studied. The BUA and VOS data were adjusted for age, age^2, height and weight. We undertook single variant tests using SNPTEST. Variants with a MAF <0.005 or not in Hardy-Weinberg equilibrium were excluded. Significant and suggestive thresholds were set at 5E-8 and 5E-6 respectively.

Results: In the single variant analysis we identified 2 variants significantly associated with BUA at genome-wide significance and a further 76 showing suggestive association. No variants were associated with VOS at genome-wide significant level, however we observed 94 variants showing suggestive association with the phenotype. For VOS, 9 of the variants showing suggestive association were located in the FSHR gene region which has previously been associated with bone mineral density (BMD) and QUS parameters in postmenopausal women. For BUA, both variants associated at genome-wide significance and 28 of those showing suggestive association were located in the WNT16 gene region which has previously been associated with BMD, cortical bone thickness and osteoporotic fracture. Of the remaining loci identified in the study as provisionally associated, most were located in novel genomic regions not previously implicated in bone metabolism.

Conclusion: In this study of whole genome sequence data for association with QUS traits, 2 variants were identified as associated at genome-wide significance with a further 170 showing suggestive association. We were able to confirm the association of variants in the WNT16 gene with bone structural phenotypes and generated supporting evidence suggesting a role for the FSHR gene region in bone metabolism. We also highlighted many other novel loci and genes for further study. These variants will undergo additional replication through in silico analyses and de-novo genotyping.

*Benjamin H. Mullin contributed equally to this work.
Targeted Sequencing, Augmented with Public Resources, Identifies a Rare C3 Allele Associated with Large Risk of Age-related Macular Degeneration

X. Zhan, D. Larson, R. Fulton, C. Wang, D. Stambolian, E. Chew, A. Swaroop, G. Abecasis

1) Biostatistics, University of Michigan, Ann Arbor, MI, United States; 2) Center of Statistical Genetics, University of Michigan, Ann Arbor, MI, United States; 3) The Genome Institute, Washington University, St. Louis, MO, United States; 4) Department of Genetics, Washington University, St. Louis, MO, United States; 5) Department of Biostatistics, Harvard University, Boston, MA, United States; 6) Department of Ophthalmology and Human Genetics, University of Pennsylvania, Philadelphia, PA, United States; 7) National Eye Institute, Bethesda, MD, United States.

Macular degeneration is one of the most common causes of incurable blindness. Common alleles in >19 loci have now been associated with disease. We set out to investigate whether rare variants in the same locus were also associated with disease risk and to compare the relative effect sizes of common and rare variants.

In collaboration with the Genome Institute at Washington University in St. Louis, we designed a sequencing study focused on 8 of the known AMD risk loci (CFH, ARMS2, C3, C2/CFB, CFI, CETP, LIPC, and TIMP3/SYN3) and 2 other candidate regions (LPL and ABCA1). We resequenced these regions in 3,124 individuals (2,335 cases and 789 controls) to an average depth of 85x. To augment the number of controls available for analysis, we designed an algorithm to identify previously sequenced samples with good coverage of our regions of interest and similar genetic ancestry. Finally, we investigated association between genetic variation in each locus and risk of disease using both single variant and gene level burden tests. We estimated the ancestry of our sequenced samples and of samples from the NHLBI exome sequencing project, identifying an ancestry matched set of disease. We set out to investigate whether rare variants in the same loci were also associated with disease risk and to compare the relative effect sizes of common and rare variants.

Using this dataset, we are investigating the genetic architecture of lipids metabolism traits. Most variants were rare, with ~180,000 (83.5%) variants having a frequency of <1%. For analysis, we developed and implemented a method that uses summary statistics measuring association at each variant and their covariance matrix to reconstruct single variant and gene-level association statistics across the entire sample. At the time of submitting the abstract, we completed first round of association analysis. Overall, 154 coding variants (139 common with MAF>1%, 15 rare with MAF < 1%) were associated with one or more lipid traits at p < 3x10^-7. In addition, 25 genes showed significant association with one or more lipid traits at p < 2.5x10^-6 using at least one of the three gene level association tests we considered (a simple burden test for variants with frequency < 5%, a variable threshold association test, and sequence kernel association test for variants with frequency < 5%). Loci with evidence for association include known and novel loci, such as ABCA1, APOC3, APOE, LDLR, PCSK9, CD300LG, CELSR2, APOH, and NPC1L1. Using this dataset, we are investigating the genetic architecture of lipids traits, to estimate the number and characteristics of rare variants contributing to plasma lipid levels, and quantify the impact of performing deep sequencing and enlarging sample sizes on the power for detecting rare variant associations.
1113W
Exome-wide association study of fetal hemoglobin levels in African Americans with sickle cell disease. S. Lessard1,2, G. LeBret1,2,1. 1) Montreal Heart Institute, Montréal, Québec, Canada; 2) Université de Montréal, Montréal, Québec, Canada.

INTRODUCTION: Reactivation of fetal hemoglobin (HbF) production in patients suffering from sickle cell disease (SCD) is a promising treatment. To develop such therapy, it is however necessary to understand the genetic mechanisms underlying the regulation of HbF levels. Genome-wide association studies have identified common variants associated with HbF at three loci (BCL11A, HBS1L-MYB and HBB), which altogether explain half of the heritability of this trait. Rare variants and low-frequency variants are expected to account for a substantial fraction of the unexplained heritability and are not tested using GWAS methodologies. METHODS & RESULTS: To address this issue, 250,000 DNA sequence variants, mostly coding and rare or of low-frequency, were genotyped in 1,495 African Americans with SCD from the Cooperative Study of Sickle Cell Disease (CSSCD) using the Illumina ExomeChip array. After quality control, 132,419 polymorphic markers and 1,301 individuals remained, with genotyping rate >99.9%. Variants with a minor allele frequency (MAF) >0.05 were analysed by linear regression against HbF using the software PLINK, accounting for age, sex and the first 10 principal components. The analysis replicated the association of HbF at the BCL11A, HBS1L-MYB and HBB loci. Excluding these loci, the lowest P-values were found at two SNPs near the KLF1 gene (P=9.08×10−4 and 1.75×10−5), an erythroid transcription factor implicated in the globin switch and hereditary persistence of fetal hemoglobin. We analyzed missense, nonsense and splice site markers with MAF <0.05 using the Optimal Sequence Kernel Association Test (SKAT-O), but no genes reached exome-wide significance (all P>2.5×10−6). From the SKAT test by taking into account both common and rare variants, we found 5 significant genes including HLA-DRB1, previously identified in a German sample. From the SKAT test, we also observed interesting differential DNA methylation patterns in PBMCs within each twin pair, indicating strong genetic and epigenetic factors for this disease, we performed whole genome, whole transcriptome, and targeted methylome enrichment and high throughput sequencing. We observed a spectrum of genomic differences within each pair of monozygotic twins discordant in asthma (as well as 1 pair discordant in allergy and 2 concordant healthy pairs), and sequenced the whole transcriptome of peripheral blood mononuclear cells (PBMCs) from 20 (including the 12 pairs for whole genome sequencing) discordant monozygotic twins pairs. In addition, we also examined DNA methylation differences between each discordant twin pair by targeted methylome enrichment and high throughput sequencing. We observed a spectrum of genomic differences within each pair of monozygotic twin, including single nucleotide variants, insertions/deletions, copy number variations and structural variations, as well as different mobile element insertions, although none seemed to be a significant contributor to the asthma phenotype. Whole transcriptome sequencing revealed surprisingly similar expression patterns in PBMCs within each twin pair, indicating strong genetic determination for gene expression in monozygotic twins. Differentially expressed genes between the discordant twin pairs in pathway analysis involved in immune response, which is a key feature for symptoms of asthma. In addition, we also observed interesting differential DNA methylation patterns in each monozygotic twin pair using bisulfite sequencing. Our study provides a unique perspective to evaluate genetic and epigenetic factors associated with this complex, common disease of asthma at the omics level.

1115F
Exome sequencing identifies variants in families with idiopathic scoliosis. E. Baschali1, K. Swindle, K. Calbatt, C. Wettby1, K. Gowan, K. Jones, N. Miller. 1) Orthopaedics, University of Colorado Denver Anschutz Medical Campus, Aurora, CO; 2) Biochemistry and Molecular Genetics, University of Colorado Denver Anschutz Medical Campus, Aurora, CO.

Idiopathic scoliosis (IS) is the most common disorder of spinal balance, which, when progressive, leads to significant morbidity and financial burden to individuals and families. Most cases are strongly suspected to be genetic, arguing for the development of novel approaches to finding causative genes for this disorder. As efforts to find the genetic causations of IS, including GWAS, have not led to major advances in disease understanding, we believe that causal variants may differ among families. To this end, we used exome sequencing to focus on multigenerational families affected with IS. We completed exome sequencing for 7 multigenerational families with IS, sequencing three affected individuals in each family (average 6 affected per family). Illumina HiSeq reads were mapped to the reference human genome sequence (hg19) with large-scale alignment software (GSNAP). Sequence calls for variants (single-nucleotide polymorphisms, insertions and deletions) were performed using the Broad’s Genome Analysis Toolkit (GATK). The program ANNOVAR was used to filter the variants by cross-referencing multiple genetic variation databases (e.g., dbSNP, 1000 genomes database, AVSIFT, etc.). Information was extracted about variant frequency (if previously known) and location within genes. Non-synonymous changes, that cause an alternate splice site, and/or aberrant stop codon were considered for further analyses. For non-synonymous changes, variants were cross-referenced to the dbNSFP database to determine whether changes to the protein were considered tolerable or damaging. Finally, a segregation filter was applied. Filtering criteria required variants to be present in all three affected individuals in the family. This bioinformatic filtering resulted in the identification of 3 to 26 potential causative variants in each family (minor allele frequency less than 5%). No variants were also present in all families. We performed confirmatory Sanger sequencing for select variants in individuals who originally underwent exome sequencing and in additional affected and unaffected family members. These findings suggest that genes with identified variants may contribute to the scoliosis phenotype of this significant step toward understanding the etiology of this disorder. Further investigation is needed to validate novel genes for IS by sequencing the genes in an additional familial IS cohort.

1114T
Targeted sequencing identifies loci associated with sarcoidosis in African Americans. I. Adrianto1, G.B. Wiley2, A.M. Levin2, S.B. Glenn, M.C. Iannuzzi, B.A. Rybicki, P.M. Gaffney, C.G. Montgomery. 1) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Public Health Sciences, Henry Ford Health System, Detroit, MI; 3) Department of Medicine, SUNY Upstate Medical University, Syracuse, NY.

Sarcoidosis is a systemic granulomatous disease that primarily affects the lungs, but can manifest in multiple organ systems. While previous genome-wide association studies (GWAS) of both European and African American populations have identified common variants associated with sarcoidosis, rare variants (minor allele frequency<1%-5%) are hypothesized to have large effect sizes and may contribute to sarcoidosis risk. Therefore, we analyzed both common and rare variants to identify loci associated with sarcoidosis. We performed targeted (Illumina HiSeq) whole exome sequencing on 4,22×103 individuals of African ancestry with sarcoidosis. Variants with a minor allele frequency (MAF) ³1,301 individuals remained, with genotyping rate >99.9%. Variants with a minor allele frequency (MAF) >0.05 were analysed by linear regression against HbF using the software PLINK, accounting for age, sex and the first 10 principal components. The analysis replicated the association of HbF at the BCL11A, HBS1L-MYB and HBB loci. From the SKAT-O test in the association results, we performed gene-set enrichment analyses using gene-based association results. Preliminary analysis revealed multiple pathways nominally enriched, implicating among others Notch, FGF and SAPK/JNK signalling. CONCLUSION: We genotyped SCD patients with the Exome Chip to identify functional DNA sequence variants associated with HbF levels. The data generated here will be combined with other genetic results along with comprehensive functional data in order to identify new regulators of HbF production and biomarkers of SCD morbidity.

1116W
Asthma: An Omics View through Discordant Monozygotic Twins. R. Chen1, G.I. Mias1, S. Runyon2, J. Li-Pook-Than1, G. Euskirchen, P. Lacroute, K.C. Nadeau, M. Snyder1. 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Pediatrics, Stanford University, Stanford, CA.

Asthma is a complex disease that has a clear hereditary component. As a common disease, asthma affects 11% of the population in the United States, or 13% in children under age 18. To determine the contribution of genetic and epigenetic factors for this disease, we performed whole genome, whole transcriptome, and targeted methylome profiling in 20 pairs of monozygotic twins, who are discordant in asthma, using high throughput sequencing. We determined the whole genomes of 12 pairs of monozygotic twins discordant in asthma (as well as 1 pair discordant in allergy and 2 concordant healthy pairs), and sequenced the whole transcriptome of peripheral blood mononuclear cells (PBMCs) from 20 (including the 12 pairs for whole genome sequencing) discordant monozygotic twins pairs. In addition, we also examined DNA methylation differences between each discordant twin pair by targeted methylome enrichment and high throughput sequencing. We observed a spectrum of genomic differences within each pair of monozygotic twin, including single nucleotide variants, insertions/deletions, copy number variations and structural variations, as well as different mobile element insertions, although none seemed to be a significant contributor to the asthma phenotype. Whole transcriptome sequencing revealed surprisingly similar expression patterns in PBMCs within each twin pair, indicating strong genetic determination for gene expression in monozygotic twins. Differentially expressed genes between the discordant twin pairs in pathway analysis involved in immune response, which is a key feature for symptoms of asthma. In addition, we also observed interesting differential DNA methylation patterns in each monozygotic twin pair using bisulfite sequencing. Our study provides a unique perspective to evaluate genetic and epigenetic factors associated with this complex, common disease of asthma at the omics level.
Whole genome sequencing identifies genes and non-coding regions as modifiers of chronic Pseudomonas aeruginosa infection in cystic fibrosis. J.X. Chong1, M.J. Emond2, T. Louie2, R.L. Gibson1, M.J. Bamshad1,2, NHLBI GO Exome Sequencing Project. 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA.

We recently used exome sequencing of phenotype extremes to identify variants in DCTN4, CAV2, and TMGC which are associated with risk for chronic Pseudomonas aeruginosa (Pa) lung infection in individuals with cystic fibrosis (CF). To investigate the contributions of coding variants not captured by existing exome targets and non-coding variation on age of onset of chronic Pa infection, we performed whole genome sequencing on individuals with CF and early-onset (n=44) or late-onset (n=45) chronic Pa. We used SKAT-O with small sample adjustment to test for differences between extremes in the distributions of variants in genes and non-coding elements, down-weighting common variants. In the per-gene analysis, we analyzed non-synonymous, splice, and utr variants. The most significant (naive p = 6.2 × 10^-7) association was observed with CD22. In the non-coding analysis, we used genome segmentation annotations from the ENCODE project to predict regulatory elements such as enhancers and insulators. We identified 44,749 predicted enhancers that were variable in at least five subjects. While current computational methods cannot reliably predict enhancer targets, the top two associations were a region upstream of ADAM19 and an intronic region in STAP1. These genes play roles in host immunity and are thus biologically plausible candidates. Additionally, we used GREAT (Genomic Regions Enrichment of Annotations Tool) to assign enhancer candidates to potential target genes and identified 10.7-fold enrichments (FDR q = 6.1×10^-4) for inositol phosphate kinase activity among the top 5% of enhancer associations. This signature is attributable to eight predicted enhancers in ITPK1 and three in/near ITPKB (a kinase upstream of ITPK1 in the inositol phosphate signaling pathway), yet neither gene was significant nor highly ranked by p-value in the per-gene analysis: ITPK1 (naive p = 0.35) was ranked 6,158th and ITPKB (naive p = 0.15) 3,041st out of 15,800 genes. Interestingly, variability in ITPK1 expression has been previously predicted to affect chloride channel activity, and two variants identified in our study are located within genes of alternative chloride channel activity, suggesting that variants in ITPK1 enhancers could lead to differences in age of onset of chronic Pa infection. Nevertheless, validation in a replication cohort is a necessary next step. This analysis represents an early attempt to associate non-coding variation discovered by WGS using an extreme phenotype study design.
1120T
Whole-genome sequencing of an Italian Multiple Sclerosis multiplex family identifies a novel functional variant in GRAMD1B, F. Martineili Boneschi1,2, F. Esposito1,2, A. Osiceanu1, A. Zauli1, D. Cittaro1, M. Sorosina1, A. Calabria1, D. Lazarevic1, V. Maselli1, P. Brambilla1, G. Comi1,2, E. Stupka2, 1) DEPT NEUROLOGY & INSPE, SCIENTIFIC INST SAN RAFFAELE, MILAN, MI, Italy; 2) Laboratory of genetics of neurological complex disorders, IINN, Scientific Institute San Raffaele, Milan, Italy; 3) Genome Function Unit & Center of Translational Genomics and bioinformatics, Scientific Institute San Raffaele, Milan, Italy; 4) San Raffaele Telethon Institute for Gene Therapy (HSR-TIGET), Scientific Institute San Raffaele, Milan, Italy.

While the role of common genetic variants is clearly established from recent studies on individual with multiple sclerosis (MS), the contribution of rare variants to the disease susceptibility in multiplex families remains unclear. Our study is to identify rare genetic variants related to MS susceptibility in an Italian multiplex family. SNP microarray genotyping and whole-genome sequencing in 4 MS patients and 4 unaffected individuals belonging to an Italian multiplex family descending from a first cousin marriage were performed. The Merlin software was used for the linkage analyses and SNPeff software was applied to prioritize rare variants. We identified 12,938 variants with high or moderate functional impact according to SNPeff software annotation; 5,453 of them are present with at least one copy in all the patients, of which 250 are novel (not present in dbSNP 137) or rare (GMAF <0.02). Six of these 250 variants are located under one of the two LOD peaks (LOD>1.5) identified on chromosome 8p21.2 and 11q23.3. Four of them are homozygous in cases and controls, while the remaining two suggested an autosomal recessive mode of inheritance with incomplete penetrance. The first one (rs201666255) maps to a region that is a olfactory receptor gene under positive selection (GERP<sup>++</sup> score=3.33), while the second one falls within the GRAMD1B gene, causing an amino acid substitution (S601P) (GERP<sup>++</sup> score=4.72). The second variant is a non-synonymous missense variant with high impact on AMNR, a gene that encodes a membrane protein, which is part of the GRAM containing domain family protein. In the mouse it is highly expressed in the central nervous system and in specific immune cell subtypes, like dendritic cells and neutrophils. Transcription experiments and functional studies are ongoing to investigate the role of the gene on the disease.

1121F
Comprehensive genomic profiling of 66 cardiometabolic phenotypes by whole genome sequencing in 3,621 samples from the UK10K project.

J. Min, The UK10K Consortium Cohorts Group. School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom.

As part of UK10K Consortium Cohorts Group, 3,621 individuals from two deeply phenotyped population-based collections - TwinsUK and the Avon Longitudinal Study of Parents and Children (ALSPAC) - have been whole genome sequenced (WGS) to average 6.5x coverage using next-generation sequencing technology. Variants discovered through WGS of the TwinsUK and ALSPAC cohorts along with those known from 1000 Genomes were imputed into the full genomewide association study genotyped cohorts increasing the sample size for single point association analysis to 12,724 subjects. In a meta-analysis of directly sequenced and imputed variants, we tested rare variants association with 66 cardiometabolic phenotypes. In single point analysis, we found 30 novel loci with P<5e-8 of which 13 with an effect allele frequency below 5%. In addition, novel low frequency trait associations underlying known loci such as LDLR and LPL and total cholesterol were found. In WGS data only, we used SKAT to analyse 26,367 gene regions using 41,692 non-overlapping windows with a maximum of 50 coding variants with MAF<5%. We identified 18 regions below a gene-based significance threshold of P<1.9e-6 including known associations such as ADIPOR2 and Adiponectin (P=3.7e-11) and PCSK9 and LDL (P=1.8e-7). Novel loci are being validated and replicated. Among the single point associations with common variants (MAF>5%), we showed a consistent pattern of enrichment of exonic and 5'UTR variants using 100,000 permutations matched on TSS and MAF (p=1e-5, r=2.01). The enrichment of exonic variants was also seen for single point associations with MAF 1-5%. As a substantial proportion of causal alleles may exist beyond the exome, we combined genome-wide maps of functional annotations including DNase hypersensitive sites (DHS) from six ENCODE cell types to infer the pathogenicity of variants in regulatory or conserved regions. We found DHS enrichments of single point associations relative to sites not in DHS peaks, DHS hotspots or exons (p<1e-5, r=2.01). We will investigate phenotype cell type specificity by comparing enrichments of cell type specific sites to sites shared among multiple cell types. Our results provide insights into how large-scale whole-genome sequencing efforts are likely to contribute to the understanding of the genetic architecture of cardiometabolic phenotypes and may therefore significantly increase previous estimates of explained heritability.

1122W
Functional annotation combined with evolutionary principles facilitate whole genome sequence analyses of complex traits: the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium.

A.C. Morrison1, F. Yu2, J. Liu3, A. Voorman4, A.D. Johnson5,6, J. Reid7, X. Liu8, D. Muzny2, A.R. Folsom9, C.J. O'Donnell1,6, B.M. Psaty9,8, G.A. Culpus1, A. Clark1, R. Gibb1, E. Boenink1,2. The Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium.

1) Human Genetics Center, University of Texas at Houston, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 3) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 4) Department of Biostatistics, University of Washington, Seattle, WA; 5) NHLBI Framingham Heart Study, Framingham, MA; 6) Division of Intramural Research, NHLBI, National Institutes of Health; 7) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 8) Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA; 9) Group Health Research Institute, Group Health Cooperative, Seattle, WA; 10) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Joint analysis of functional annotation of genome variation, particularly outside of non-protein coding regions, combined with empirical evidence of natural selection indicating phenotypic consequences should prove to be an efficient approach to analyzing whole genome sequence data. Here we intersect the results of sequencing the whole genomes of 2,700 deeply phenotyped individuals from the CHARGE consortium, annotation using multiple predictive functional genomic resources (e.g. dbNSFP and ENCODE), and application of population genetic measures of variation (e.g. Watterson's J, integrated haplotype score (iHS)) to infer regions of natural selection. Predicted functional domains (e.g. transcription factor binding sites) with significant evidence of natural selection are used as the unit of inference for genotype-phenotype associations. Segregates within the family being homologous in 3 domains (e.g. transcription factor binding sites and microRNAs) reveal an abundance of evidence for selection acting on noncoding regions of the genome. Among the regions with the strongest evidence of natural selection, ~60% lie in intergenic regions, ~33% in intronic regions, and slightly over 1% in genic regions. Focused analyses stratified by functional classes of noncoding variants detects pervasive purifying selection acting on enhancers, transcription factor binding sites, microRNAs and target sites but not on lincRNA or piRNA. Integration of functional annotation and whole genome sequence data reveals that natural selection affecting non-coding regions of the genome influencing health and disease.

9) Group Health Research Institute, Group Health Cooperative, Seattle, WA; 10) Department of Biostatistics, Boston University School of Public Health, Boston, MA.
1123T

Exome sequence variants associated with blood levels of hemostatic factors: The Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium and the NHLBI Exome Sequencing Project (ESP). N. Pankratz1, J. Brody2, M.H. Chen3, B. Davis4, P. Wei5, J. Reid6, C.J. O’Donnell6, N.L. Smith7, A.C. Morrison8 on behalf of the CHARGE Hemostasis Working Group and the ESP Hemostasis Working Group. 1) University of Minnesota, Minneapolis, MN; 2) University of Washington, Seattle, WA; 3) Boston University, Boston, MA; 4) University of Texas Health Science Center at Houston, Houston, TX; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) NHLBI’s Framingham Heart Study, Framingham, MA.

Perturbed hemostasis, reflected by altered levels of blood coagulation proteins, can lead to morbidity and mortality from arterial and venous thrombosis. Hemostatic factors such as fibrinogen, coagulation factors VII and VIII, and von Willebrand factor (vWF), are intermediate phenotypes that are associated with clinical thrombotic events; thus, it is important to study these traits and identify genetic variants that increase susceptibility to clotting. To identify associations with rare and low frequency variants, we performed exome sequencing in independent samples from CHARGE and ESP. Sequencing in CHARGE was performed in a subset of participants from the Atherosclerosis Risk in Communities Study (ARIC; n=3,173 whites, 2,408 blacks), the Cardiovascular Health Study (CHS; n=741 whites), and the Framingham Heart Study (FHS; n=850 whites). ESP cohorts with available hemostasis measures included ARIC, CHS, FHS, Coronary Artery Risk Development in Young Adults (CARDIA), Multi-Ethnic Study of Atherosclerosis (MESA), and Women’s Health Initiative (WHI), for a total of 2,649 EAS and 1,925 AAS. Available measures of fibrinogen, factor VII, factor VIII, and vWF varied across studies and projects. Initial results (n=3,851 whites total) are described for the meta-analysis of ESP and an initial subset of the CHARGE data for these four hemostatic factors using three approaches: 1) single variant test using linear regression for all variants above 5% minor allele frequency (MAF); 2) a T5 test where all non-synonymous or splice-site variants with a MAF of 5% or lower are summed together to generate a gene score; and 3) the SKAT test which analyzes the same variants as the T5 test, but allows for effects to be in either direction and upweights the contributions from common variants. The gene-based tests will be fully evaluated in the total combined sample of 6,225 whites and 3,908 blacks across ESP and CHARGE. These efforts represent the largest integration of exome sequences from two large national projects to identify novel genes and variants associated with inter-individual variation in hemostatic factors.

1124F

Exome sequencing study identifies several candidate variants associated with Kawasaki disease. Y. Park, J. Kim, J. Lee, Korean Kawasaki Disease Genetics Consortium, Asan Institute for Life Sciences, University of Ulsan College of Medicine, Seoul, Korea.

Kawasaki disease (KD) is an acute systemic vasculitis of infants and children, manifested by fever and signs of mucocutaneous inflammation. The highly effective and standard treatment for KD is high-dose intravenous immunoglobulin (IVIG) therapy. However, about 20 % of KD patients have persistent or recurrent fever after the initial IVIG treatment, which increases the risk for coronary artery lesions. To identify susceptibility variants for KD, we sequenced whole exomes of 12 patients with KD as well as 17 control exomes. As unique variants to KD compared to control exomes, a total of 43 novel nonsynonymous single-nucleotide polymorphisms (nsSNPs) were detected by exome sequencing and validated by capillary resequencing. To replicate the findings, additional 18 KD patients and 18 healthy controls were analyzed by direct sequencing. We examined the associations between 30 KD patients and 35 controls in the combined data (exome sequencing data and small scale of case-control data) analysis, and identified strong signal of association with KD in 8 loci. Among 8 loci, 4 genes (FDE3A, TNFSF8, IL16 and CLEC11) had variants unique in KD samples only and the other 4 genes (HDA-DQ8, SNAPC4, MLL and HAVCR2) showed high odds ratio (OR > 4.9). Currently we are investigating whether these 8 nsSNPs are significantly associated with Kawasaki disease in large independent samples of KD case and control. Our initial data suggest that whole exome sequencing can be used for the identification of causal variants of KD.
1126T

Rare functional variants in complement genes mitigate C3 inactivation and confer high risk of advanced age-related macular degeneration. S. Raychaudhuri1,2,4,5, Y. Yu6, E.C. Miller4, R.M. Reynolds6, P.L. Tan7, S. Gowrisankar3, J.T. Goldstein3,9, M. Triebwasser10, H.E. Anderson10,11, J. Zerbib11, D. Kavanagh11, E. Soulié11, N. Katsanis6, M.J. Daly3,4, J. Atkinson3, J.M. Seddon4,12, 1) Division of Genetics, Brigham and Women’s Hospital, Boston, MA, USA; 2) Division of Immunology, Allergy, Brigham and Women’s Hospital, Boston, Massachusetts, USA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 4) Partners HealthCare Center for Personalized Genetic Medicine, Boston, Massachusetts, USA; 5) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK; 6) Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Boston, Massachusetts, USA; 7) Division of Rheumatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; 8) Center for Human Disease Modeling and Departments of Cell Biology and Pediatrics, Duke University, Durham, North Carolina, USA; 9) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 10) Institute of Genetic Medicine, Newcastle University, International Centre for Life, Newcastle upon Tyne, UK; 11) Hôpital Intercommunal de Créteil, Hôpital Henri Mondor, Université Paris Est Créteil, Créteil, France; 12) Department of Ophthalmology, Tufts University School of Medicine, Boston, Massachusetts, USA.

Motivation: We sought to identify rare coding variants which influence risk of advanced age-related macular degeneration (AMD), in order to better define disease mechanism. We were motivated by the discovery of the R1210C mutation in CFH by our group. This mutation confers high risk of AMD (OR=20) and causes loss of C-terminal function.

Methods: To find additional rare coding functional variants we applied next-generation sequencing to target the exons of 681 genes within AMD-associated loci and pathways in 2,483 cases and controls. We only identified one gene, CFI, with a significant burden of rare variants in cases. We observed that 7.8% of AMD cases compared to 2.3% of controls are carriers of rare missense CFI variants in our data (OR=3.6, p=2×10⁻⁸). Moreover, analyses of these data correctly recognised four known disease-conferring rare missense CFI variants, including mutations resulting in likely haploinsufficiency.

Conclusions: The combination of functional studies, and rare variants of complement genes outside CFI, define disease mechanism. We were motivated by the discovery of the R1210C mutation in CFH by our group. This mutation confers high risk of advanced age-related macular degeneration.

1127F

Exome sequencing in Autism Spectrum Disorder. S. Walker1, R. Yuen1, B. Thruvahindrapuram4, L. Lau1, C.R. Marshall4, B. Fernandez2, E. Fombonne3, W. Roberts1, Zwaigenbaum1, P. Szatmari1, S.W. Scherer1, 1) Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Memorial University of Newfoundland, Disciplines of Genetics and Medicine, St. John’s, A1B 3V6, Canada; 3) Montreal Children’s Hospital and McGill University, Department of Psychiatry; Montreal, H3Z 1P2, Canada; 4) The Hospital for Sick Children, Autism Research Unit, Toronto, M5G 1L7, Canada; 5) University of Alberta, Department of Pediatrics, Edmonton, Canada; 6) McMaster University, Department of Psychiatry and Behavioral Neurosciences, Hamilton, L8S 4K1, Canada.

To discover rare genetic variants and new genes associated with Autism Spectrum Disorder (ASD), we are performing a detailed genomic analysis in a cohort of 1000 Canadian Families. By combining high resolution microarrays with exome and whole genome sequencing, we aim to investigate the role of de novo and rare inherited mutations, incorporating copy number (CNV), single nucleotide (SNV) and insertion/deletion variants.

From the first 700 individuals analysed by exome sequencing with SOLiD 5500xl, we found in the order of 24,000 single nucleotide variants per individual, of which approximately 350 are novel and in coding regions. We have uncovered numerous rare and novel inherited variants that appear to segregate with the phenotype in genes previously associated with ASD such as NRXN1, NLGN4X, ARID1B and CHD8, including mutations resulting in likely haploinsufficiency. De novo mutations have also been discovered implicating new genes such as RIMS2, NTF3 and LYPD6B in the disorders. We have found multiple individuals carrying potentially pathogenic both CNVs and SNVs and in some cases more than one arising de novo. Additional incidental findings of rare coding SNVs were identified in genes KIAA1277, FAT3, STXBP5L and a single base insertion in USP54.

Our data support a multigenic, multifactorial model for Autism susceptibility and highlight the necessity for extensive information of both genotypes and phenotypes to further our understanding of complex disorders.

1128W

Exome sequencing in schizophrenia quartets families identifies the patterns and rates of causal mutations and CNVs. D. Zhou1, Z. Zhang1, Y. Liu1, L. He1,2, L. Zwaigenbaum3, L. Lau1, C.R. Marshall4, P. Szatmari1, E. Fombonne3, R. Yuen1, 1) Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Bio-X Institutes, Shanghai Jiao Tong University, Shanghai 200030, People’s Republic of China; 2) Institutes of Biomedical Sciences, Fudan University, 303 Mingdao Building, 138 Yixueyuan Road, Shanghai 200032, People’s Republic of China.

Schizophrenia is known to be a multifactorial disorder with an underlying heritability that is believed to be around 80%, without totally displaying Mendelian inheritance. Previous studies have commonly thought that many genes (or other functional genomic elements) are involved. However, the common variants and mutations in known genes confer only a limited effect on the phenotype, and these could not explain a substantial fraction of schizophrenia heritability. Here we examined the effects of the novel rare mutations, CNVs on schizophrenia pathogenesis by sequencing the whole exomes of about 300 subjects from 50 schizophrenia quartets families (only one cases in parents, more than two cases in children). We identified the rare mutations or large copy number alternations in each family. We will then use protein-protein interaction (PPI) network to evaluate the connection of the causal genes.
Identification of low frequency variants associated with albuminuria and kidney function in Danes with and without diabetes through exome-array analysis. T.S. Ahluwalia, J. Bork-Jensen1, N. Granot1, R. Ribel-Madsen1, J.M. Justesen1, M.N. Harder1, T.H. Sparasci1, T.O. Kielplaaßen1, T. Skaarby1, C. Christensen1, I. Brandslund2,5, M. Aaddahl1, M.E. Jergensen1, A. Linneberg1, L. Husmoen1, T. Lauritzen1, T. Jergensen1, P. Pedersen1,6,11, T. Hansen1,6,11, O. Pedersen1,6,11, 1) Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 2) Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark; 3) Department of Internal Medicine and Endocrinology, Vejle Hospital, Vejle, Denmark; 4) Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark; 5) Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark; 6) Steno Diabetes Center, Gentofte, Denmark; 7) School of Public Health, Department of General Practice, University of Aarhus, Aarhus, Denmark; 8) Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 9) Faculty of Medicine, University of Aalborg, Aalborg, Denmark; 10) Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; 11) Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark.

Type 2 diabetes (T2D) is an increasingly common disease with 347 million people affected worldwide. About 35% of individuals with T2D attain microvascular complications of the kidney which may lead to end stage renal failure and death. Diabetic kidney disease has been shown to be partially caused by a genetic susceptibility component. We therefore sought to identify coding variants associated with kidney function through an exome-array analysis. T.S. Ahluwalia, J. Bork-Jensen1, N. Granot1, R. Ribel-Madsen1, J.M. Justesen1, M.N. Harder1, T.H. Sparasci1, T.O. Kielplaaßen1, T. Skaarby1, C. Christensen1, I. Brandslund2,5, M. Aaddahl1, M.E. Jergensen1, A. Linneberg1, L. Husmoen1, T. Lauritzen1, T. Jergensen1, P. Pedersen1,6,11, T. Hansen1,6,11, O. Pedersen1,6,11, 1) Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 2) Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark; 3) Department of Internal Medicine and Endocrinology, Vejle Hospital, Vejle, Denmark; 4) Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark; 5) Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark; 6) Steno Diabetes Center, Gentofte, Denmark; 7) School of Public Health, Department of General Practice, University of Aarhus, Aarhus, Denmark; 8) Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 9) Faculty of Medicine, University of Aalborg, Aalborg, Denmark; 10) Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; 11) Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark.

We used the exome-wide association approach by deploying the customized Human Exome Genotyping Bead Chip manufactured by Illumina; this particular array contains ~250,000 loci. Genotyping was performed in three groups: T2D, non-T2D, and one including 2,000 Danes. After filtering the variants for minor allele frequency (MAF) > 0.1% and genotyping call rate of > 0.90, the final associations were examined in each cohort. Adjustments for age, sex, and four principal components were made for the association between SNP genotypes and POAG risk using unconditional logistic regression. Further compensations were made for the top axes of genetic stratification for each independent collection to ensure minimal confounding due to differing ancestries between POAG cases and controls. Meta-analysis summarizing the results across all cases and controls was performed using fixed effects modeling weighted in an inverse-variance manner. Results: We have confirmed strong evidence of association with variants of several genes, including CDKN2BAS (OR=0.74, P = 1.27 x 10^-10) and several novel loci. Novel variants underlying POAG are being validated in an additional 4,000 cases and 11,000 controls that have been independently collected. Conclusions: We have performed a meta-analysis of a large multi-ethnic and racial dataset that has confirmed the associations between CDKN2BAS and POAG. In addition we have identified several novel genetic variants that are in the process of validation at this time.

Grant support: This work is supported by Biomedical Research Council (BMRC) in Singapore, Ref: BMRC 10/13/35/19/675.
1131W

Rare Variant Association of von Willebrand Factor Levels in a Healthy Young Cohort of European Individuals Using HumanExome BeadChip Data. A.B. Ozel1, K. Desch2, D. Siemieniak2, D. Ginsburg1,2, J. Li1, J. Rotter1, D. Bowden1, E. Spielotes3, C. Langefeld4, 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Pediatrics, University of Michigan, Ann Arbor, MI; 3) HHMI, University of Michigan, Ann Arbor, MI.

von Willebrand factor (VWF) is an abundant plasma glycoprotein that regulates hemostasis. Plasma levels of VWF are highly variable among healthy people, with 65% of its variance attributable to inherited factors. We previously performed genome-wide association and linkage analyses of VWF in a healthy sibling cohort of 1,152 subjects and a second healthy cohort of 2,310 individuals (Desch et al., 2012). VWF plasma levels showed significant association with variants at the ABO locus (p<7.9E-139) and VWF locus (p<5.5E-16), consistent with previous reports. Linkage analysis using sibling data identified significant signals at a ~35cM interval on 2q12-2p13 (LOD=5.3), and a ~10cM interval on 8q34 at the ABO locus (LOD=2.9), which explained 19.2% and 24.5% of the variance in VWF levels, respectively. The presence of a strong linkage signal and its absence in association studies suggest a causal locus harboring many genetic variants that are individually rare, but more common in aggregate. To investigate the role of rare variants (MAF<=5%) we collected genotype data on 940 European individuals in the first cohort using the Illumina HumanExome Beadchips. Most of the >240K variants were rare coding variants, with an average call rate of 99.7% after data clean-up. Common variants on the exome chip replicate previous findings of significant association at ABO (p<1E50). Preliminary analysis of 211,149 low-frequency variants (MAF<=5%, HWE<1E-6) in 15,960 genes using the Sequence Kernel Association Test optimal (SKAT-O) gene-burden test showed a gene-wise significant association for 20 rare variants at SLFN1 (Schlafen-Like 1) on chromosome 1 (p<2E-6). Schlafen-family genes may have regulatory roles in haematopoiesis due to their increased expression levels during myeloid differentiation. The top signal in the chromosome 2 linkage interval was for 21 variants at the RETSAT gene (Retinol Saturase) (p<8E-3), which plays a role in metabolic adaptation and body weight maintenance. Further tests involving other gene-burden tests (e.g., Variable-Threshold (VT), the method by Madsen & Brown-extend the current findings.}

1132T

Exome Chip-based Association Analysis Identifies Novel Coding Variants Associated with Adiposity Traits in Hispanic Americans: The IRAS Family Stud. N. Passari1, C. Gao2, J. Norris2, C. Langefeld7, C. Gao2, J. Norris2, 1) Biochemistry Dept, Wake Forest School of Medicine, Winston-Salem, NC; 2) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 3) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Epidemiology, Colorado School of Public Health, Aurora, CO; 6) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 7) Division of Gastroenterology and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI.

Despite recent successes of genome-wide association studies (GWAS) in identifying genetic variants that affect adiposity traits, a substantial proportion of the heritability remains unexplained. A potential role for low-frequency and rare variants has been proposed. Exome chip genotyping, a cost-effective approach to evaluate coding variation, was performed in 1,414 Hispanic Americans from the Insulin Resistance Atherosclerosis Family Study (IRASFS). Evaluation of 81,560 autosomal variants for the full complement of adiposity-related traits, including body mass index (BMI), waist circumference, waist-to-hip ratio (WHR), computed tomography (CT) measures of visceral (VAT) and subcutaneous (SAT) adipose tissue, and DEXA measured percent body fat, we identified two loci at exome-wide significance (P<0.05/81,560=6.13E-07). A novel coding variant in IDH1 (rs34218846, V170L, 7% MAF; P=2.51E-07-2.14E-08) obtained genome-wide significance (P<5.0E-08) for BMI, waist and WHR. This gene was further supported by complementary GWAS and imputation data (P=0.019-4.87E-06). IDH1 encodes cytosolic isocitrate dehydrogenase 1 and is postulated to play an important role in fat and cholesterol biosynthesis as demonstrated in cell and knock-out models. In addition, a rare missense variant in ALDH1A3 (rs1130737, R15G, 0.01% MAF; P=2.39E-07) was associated with WHR. Although aldehyde dehydrogenase 1 has been examined for its role in sex-differences, these results highlight that examining exonic variation could provide new insight into the regulation of hemostasis through VWF and reveal novel loci of extended heterogeneity that may explain a significant portion of the missing heritability for VWF.
Whole Exome Sequencing (WES) has proved to be a powerful method in studying Mendelian diseases. However, the identification of causative variants remains challenging due to the very large number of variants observed in each individual's exome. To narrow down the original set of variants to a short list of candidate causal mutations, a large number of tools have been developed, most of which are designed for annotating, filtering and prioritizing variants using allele frequencies in control/disease populations and computational predictions of deleteriousness and conservation of sequence variants and indels. Multiple methods are capable of utilizing trio information to detect de novo variants and/or exclude mutations violating expected Mendelian segregation. However, few published and freely available tools address the problem of selection/prioritization of variants that co-segregate in a family with a Mendelian condition under pre-specified inheritance model. All of them have significant limitations, such as low specificity or unavailability of source code. Here, we present an open source R package for co-segregation analysis that incorporates genotype data from WES, pedigree structure and phenotypic information about individuals. In the first step, the algorithm filters out all common variants and generates a 'project level VCF' based on the genotype data from all family members, i.e. the file with the genotype calls across all samples in pedigree at every variant position called in at least one sample. Next, for each variant, the pattern of co-segregation of genotype and phenotype information is evaluated under different inheritance models. At the same time, variants detected to be like, i.e. different inheritance models, are filtered based on variant frequency, and the list of variants observed to be co-segregating with disease can then be filtered/prioritized based on other criteria, such as predicted functional effects, relevant tissue expression and a priori candidate gene. Our algorithm pays particular attention to a few low quality genotype calls by removing from further filtering candidates in affected individuals and less in checking the presence of the same variants in unaffected ones. The proposed algorithm is context dependent and can 'learn' from early samples and then applied to newly recruited samples. We are successfully applying this algorithm to data from NHGRI's Centers for Mendelian Genetics, and, once published, the code will be freely available.

Identification of de novo variants contributing to nonsyndromic cleft lip and palate. E.J. Leslie1,2, K.M. Steinberg1, D.C. Koldobt1, C. Harris2, D.E. Larson3, R.S. Fulton4, G.L. Wehby3, J.T. Hecht1, T.H. Beaty5, A. Scott1, M.L. Marazita4, G.W. Weinstock5, J.C. Murray2. 1) Dept. Pediatrics, University of Pennsylvania, Philadelphia, PA; 2) Dept. of Biostatistics, Brigham & Women's Hospital, Boston, MA; 3) Dept. of Biostatistics, Harvard School of Public Health, Boston, MA; 4) Center for Medical Informatics, Harvard Medical School, Boston, MA; 5) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA. Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a common birth defect, affecting 1/1000 individuals worldwide. The complex inheritance of NSCL/P reflects the combined action of multiple genetic and environmental risk factors. Efforts to identify genetic risk factors have included numerous linkage, candidate gene, and genome-wide association (GWA) studies. To date, there are four independent GWA studies and a meta-analysis which have collectively identified genome-wide significant associations for at least a dozen loci. Despite this progress, identifying specific, etiologic variants remains a challenge. We selected thirteen regions from previous GWA and candidate gene association studies for targeted capture and deep sequencing to identify common and rare etiologic variants for NSCL/P. We sequenced these regions, totaling 6.7Mb in 1200 case-parent Asian trios from the Philippines and China plus 400 trios of European descent. Sequencing trios affords us the opportunity to identify de novo mutations which may be more deleterious than other rare variants and could contribute to the missing heritability in complex traits. De novo variants were called using the software tool Polymutt, which considered relationships between subjects. We identified 135 de novo variants absent from dbSNP and the NHLBI Exome Sequencing Project databases. Of these, 32 individuals had at least one variant in common. In particular, de novo variants in PAX7, ABCA4, and PIK3R5 and a splice site variant in IRF6. Of the remaining noncoding variants, 51 occurred in putative regulatory elements defined by histone modifications, transcription factor ChIP-Seq, and DNase hypersensitivity clusters. Notably, one individual had two rare missense de novo variants in PAX7, ABCA4, and PIK3R5 and a splice site variant in IRF6. The remaining de novo variants are likely contributors to the etiology of NSCL/P. Future analysis of these variants may facilitate identification of inherited rare and common causative variants at these targeted regions.
1137W

Exome sequencing to identify de novo and rare recessive mutations in sporadic ALS, K. Meltz Steinberg1, D.C. Koboldt1, D.E. Larson1, G.E. Sanders1, R. Pamphlett2, E.R. Mardis1. 1) The Genome Institute, Washington University, St. Louis, MO; 2) The Stacey Motor Neuron Disease Laboratory, Department of Pathology, The University of Sydney, Sydney, New South Wales, Australia.

Amyotrophic lateral sclerosis (ALS) is characterized by progressive motor neuron degeneration. Five to ten percent of patients have a family history of ALS (FALS), but most cases are sporadic (SALS). While the almost two thirds of FALS cases can be explained by mutations in a few genes, such as SOD1, the causes of SALS are still largely unknown. The late onset of disease further complicates efforts to elucidate the genetic causes of SALS using traditional family based studies. To further investigate the genetic contribution to SALS we performed exome capture and sequencing on 45 trios with no family history of ALS. The average age of the patients at onset was 47.5, and two patients were positive for hexanucleotide repeat expansion of C9orf72, which is known to contribute to the ALS phenotype. Using the trio-aware genotype caller, Polymutt, we identified 22 high confidence de novo coding mutations, and twelve of these were identified as deleterious or damaging using bioinformatic prediction software. We also identified a splice site mutation and one nonsense mutation. Of note, some of the de novo mutations characterized are in genes involved in axonal guidance and apoptosis. In addition, we identified a de novo start codon mutation in CHRNA1, a gene that harbors another de novo nonsense variant from a recently published, independent SALS cohort. Leveraging the trio information, we were also able to identify rare (MAF<0.01) compound heterozygous mutations in 18 different genes that are predicted to be highly damaging and cause a loss of function. We also identified five rare homozygous stop gain mutations and three rare homozygous splice donor/acceptor mutations. These de novo and rare recessive mutations may provide insight into the genetic contribution to SALS.

1138T

Novel intellectual disability genes identified by exome sequencing. R. Rabionet1, L. Domenech1, O. Drechsel1, M. Viñas1, A. Puig1, M. Gehre2, S. Ossowski3, I. Madrigal4, M. Guirt4, M. Mila5, X. Estivill1. 1) Center Genomic Regulation (CRG), UPP and CIBERSESP, Barcelona, Spain; 2) Center Genomic Regulation (CRG) and UPP, Barcelona, Spain; 3) Servei de Genètica, Hospital Clínic de Barcelona, Barcelona, Spain; 4) Laboratori Genètica, UDIAT-Centre Diagnostic, Corporació Sanitària Parc Taulí, Sabadell, Spain.

Intellectual disability (ID) is a genetically heterogeneous disorder affecting 1-3% of the population. About 30% of cases of ID can be explained by structural variants, and more than 100 genes have been implicated in its pathophysiology. Nevertheless, a large proportion of cases remains unexplained. A recent study has shown that de novo point mutations are a frequent cause of ID, making next generation sequencing technologies very powerful tools to identify rare de novo genetic causes of ID. We recruited 33 trio cases and 6 sibling pairs affected with ID in the ID Unit of the clinic, negative for the fragile X expansion and without cytogenetically visible abnormalities. Exome sequencing was performed, and non-synonymous variants present in the cases were selected and filtered based on their frequency in known de novo mutation databases in the trio-aware genotype caller, Polymutt. We identified five rare homozygous or compound heterozygous variants in the cases, of which an average of 27 were considered de novo. Frequency and functionality based filtering reduced the number of potential candidate ID genes harboring de novo variants to 0-10 per case. The presence and functionality based filtering reduced the number of potential candidate ID genes harboring de novo variants to 0-10 per case. The variants’ functionality score was predicted using Illumina HumanExome Beadchip array and exome array data. We performed targeted exon resequencing analysis of 770 genes proximal to CVAS regions in 2495 ID patients and 1,121 control subjects. We also used the illuminina HumanExome Beadchip array in 3,352 subjects (895 ID, 2,447 controls) to test association of coding variants with ID. In the exome-chip data we did not find any genes passing exome-wide significance threshold (P<2.5×10^-6) when testing putative loss of function (LoF) or protein altering mutations. For genes in CVAS regions we find suggestive evidence of association of putative LoF mutations at IFIHI1 (P = 0.0075) and DPP4 (P = 0.0047). Interestingly, the putative LoF mutations at IFIHI1 in the exome chip data are the same mutations reported in Nejentsev et al. 2009 to confer protection to Type 1 Diabetes, whereas in ID we find evidence of conferred risk. Furthermore, we find suggestive evidence of putative LoF mutations in 8 additional genes with a putative functional role in the central nervous system.

1139F

Rare coding variant association study of Inflammatory Bowel Disease. M.A. Rivas1, T. Green2, C. Stevens2, J.H. Cho3, J.D. Rioux4, R.J. Xavier4, M.J. Daly5, NIDDK IBD Genetics Consortium. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Broad Institute, Cambridge, MA, USA; 3) Yale School of Medicine, New Haven, Connecticut, USA; 4) Université de Montréal and Research Centre, Montreal Heart Institute, Montreal, Quebec, Canada; Center for Computational & Integrative Biology, Cambridge, MA, USA; 6) Analytical and Translational Genetics Unit Massachusetts General Hospital, Boston, MA, USA.

In a prior pilot study we reported evidence of independent rare coding variants in common variant association study (CVAS) regions associated to inflammatory bowel disease (IBD). In this study we explored the extent to which rare and low frequency variants contribute to susceptibility of IBD through a combined analysis of a large exome resource and exome array data. We performed targeted exon resequencing analysis of 770 genes proximal to CVAS regions in 2495 ID patients and 1,121 control subjects. We also used the Illumina HumanExome Beadchip array in 3,352 subjects (895 ID, 2,447 controls) to test association of coding variants with IBD. In the exome-chip data we did not find any genes passing exome-wide significance threshold (P<2.5×10^-6) when testing putative loss of function (LoF) or protein altering mutations. For genes in CVAS regions we find suggestive evidence of association of putative LoF mutations at IFIHI1 (P = 0.0075) and DPP4 (P = 0.0047). Interestingly, the putative LoF mutations at IFIHI1 in the exome chip data are the same mutations reported in Nejentsev et al. 2009 to confer protection to Type 1 Diabetes, whereas in ID we find evidence of conferred risk. Furthermore, we find suggestive evidence of putative LoF mutations in 8 additional genes with a putative functional role in the central nervous system.

1140W

Exome Array Identifies Novel Loci and Rare Variants Associated With Age-related Macular Degeneration. Y. Yu1, S. Raychaudhuri2,3,4,5,6, R. Reynolds1, J.I. Goldstein1, E. Souied1, M.J. Daily1, J.M. Seddon2,5,6 1) Ophthalmic Epidemiology and Genetics Service, Department of Ophthalmology, Tufts Medical Center, Boston, MA; 2) Division of Genetics, Brigham and Women’s Hospital, Boston, Massachusetts, USA; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women’s Hospital, Boston, Massachusetts, USA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 5) Partners HealthCare Center for Personalized Genetic Medicine, Boston, Massachusetts, USA; 6) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK; 7) Analytic and Translational Genetics Unit Massachusetts General Hospital, Boston, Massachusetts, USA; 8) Hôpital Intercommunal de Créteil, Hôpital Henri Mondor, Université Paris Est Créteil, Créteil, France; 9) Department of Ophthalmology, Tufts University School of Medicine, Boston, Massachusetts, USA.

Meta-analysis of Genome-wide association studies (GWAS) has identified 19 common variants for age-related macular degeneration (AMD), however, a large proportion of the heritability of AMD is still unexplained and the functional consequences of these variants are still unclear. To fine-map AMD loci and identify rare variants in a gene related to the inflammatory pathway for which no common variants have been associated with AMD previously. This study supports the involvement of both common and rare variants in AMD pathogenesis. Additional variants in the low-frequency spectrum are likely to explain the missing heritability of AMD.
Exome sequencing in children with severe viral respiratory infections.


Background: Human genetic variation plays a key role in determining individual responses after exposure to infectious agents. Respiratory viruses are the most common pathogens leading to hospitalization in children under the age of 5. Infections are usually mild and self-limiting, however, 0.1-0.2% of all children require admission to Intensive Care Unit (ICU). The majority of these children are previously healthy without known risk factor. Here, we use exome sequencing to search for genetic variants conferring unusual susceptibility to viral respiratory infection in the pediatric population. Methods: Study participants are recruited since 2011 in Swiss and Australian pediatric ICU. Inclusion criteria are age 0-3, ICU admission due to viral respiratory infection, mechanical ventilation, and absence of known risk factor or co-morbidity. So far, we have captured and sequenced the exomes of 103 children with acute respiratory infections in the ICU. We used VAAST for pathway analysis. Results and discussion: Samples were sequenced at high coverage (72x on average). After filtering, we observed an average of 8434 synonymous-coding, 7889 non-synonymous-coding, 71 splice-sites, 81 stop-gain and 7 stop-loss variants per sample. Using a VAAST score threshold of 30, the case-control analysis resulted in the identification of 288 genes enriched for potentially deleterious variants in cases. From this list, we filtered out 38 highly variable genes (HLA, mucins, olfactory receptors, T-cell receptors and zinc fingers) as well as an additional set of 73 genes that systematically obtained high VAAST scores in permutation analysis, resulting in a final number of 177 candidate genes. Functional classification of these candidates using DAVID showed enrichment for innate immunity genes. Our ongoing recruitment and sequencing effort will allow us to explore candidate variants, genes and pathways in a larger study population.

Testing rare coding variation for an impact on HIV-1 viral load through exome sequencing.

P.J. McLaren1,2, P.R. Shea3, D.B. Goldstein3, J. Fellay1,2, 1 School of Life Science, EPFL, Lausanne, Vaud, Switzerland; 2 Institute of Microbiology, University Hospital Lausanne, Lausanne, Vaud, Switzerland; 3 Center for Human Genome Variation, Duke University School of Medicine, Durham, NC, USA.

Background: Common variants (maf>5%) in the MHC and CCR5 regions are known to influence set point HIV-1 viral load (spVL) yet explain only a portion of the total trait variance. The impact of rare coding variation on HIV-1 disease progression has not been as thoroughly investigated. Here we utilize exome sequencing in 391 HIV-1 infected individuals with stable spVL to look for rare and functional variants that mediate control of HIV-1 infection. Methods: Set point HIV-1 viral load was calculated as the average of at least 3 measurements obtained during the chronic phase of infection. We captured and sequenced all coding exons to high coverage (>70x) in 391 HIV-1 infected individuals of the Swiss HIV Cohort Study using the Illumina Truseq 65Mb enrichment kit and the Illumina HiSeq2000. Paired-end, 120 bp reads were aligned using the Burrows-Wheeler Aligner (BWA). Quality control and variant calling were performed using GATK. Variant functional annotation was performed using snpEff version 2.1. Individual variants were tested for association using linear regression. Gene and gene network burden testing was performed using SCORE-Seq and SKAT. Results: Individual variant testing showed strong evidence for association in the MHC region. Conditional regression demonstrated that this signal is explained by two independent SNPs, rs1057151 (p=6.7e-11) in HLA-B and rs3207555 (p=4.0e-10) in HLA-C. Accounting for these two SNPs, no other variants show evidence for association. Testing for a relationship between burden of rare variants (maf<5%) and spVL in 16,839 genes individually did not uncover significant associations. Similarly, there was no evidence for enrichment of rare variants affecting spVL across a network of 2,971 genes identified as interacting with HIV-1 in biochemical screens. Conclusion: Outside of the MHC, no significant impact of rare variation on spVL was detected by exome sequencing in 391 individuals. Larger samples are likely required to fully explore the role of rare coding variation on this phenotype. Additional classes of variation not detected by GWAS or current sequencing technologies may also contribute to host HIV-1 control.
1144T Cumulative effect of coding sequence variation in TLR6 and ENG influences risk of infectious complications in patients with Staphylococcus aureus bacteremia. W.K. Scott1,2, D.M. Dykhoff1,2, S. Guo1, C.L. Nelson3, T. Rude4, F. Ruffin4, A.S. Allen5,6,7, Q. Yan4, V.G. Fowler6,7,8,1.1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami FL; 3) Duke Clinical Research Institute, Duke University Medical Center, Durham NC; 4) Department of Medicine, Duke University Medical Center, Durham NC; 5) Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham NC.

Staphylococcus aureus (SAB) is a serious, common infection. Both host and bacterial genetic variation likely influence development of complications in SAB patients. Studies of common single nucleotide variants (SNV) have failed to detect significant, reproducible associations with SAB complications. Because such studies may miss the impact of rare coding SNV due to weak linkage disequilibrium (LD), we used whole-exome sequencing to determine if the cumulative effect of common and rare coding SNV influences risk of complicated SAB. To reduce the effect of bacterial strain, we analyzed patients infected with S. aureus clonal complexes 5 or 30. In the first phase of the 200-sample study, 35 white cases of complicated SAB were compared to 35 age- (within deciles) and sex-matched white SAB controls without complications. After DNA capture with the Agilent SureSelect 72Mb exome/UTRs kit, samples were sequenced 3 per lane on the Illumina HiSeq 2000. DNA sequences were aligned with BWA and genotypes called with GATK and VQSR recalibration. SNV with VQSLOD≥2 and quality score of 99 were analyzed. Genotypes with read depth<8 and low-confidence SNV were removed before analysis. Association of genes with complicated SAB was tested by Cochran-Armitage (CA) -Sum and -Max tests using RVASSOC software. CA-Sum sums individual SNV chi-square test statistics across the gene and evaluates significance by permutation. CA-Max considers only the SNV with the largest chi-square statistic. While no gene was significantly associated with complicated SAB after Bonferroni correction (p=0.05/17,830 genes=2.8×10−5), the most significant genes (p<10−4 on both tests) were TLR6 (CA-Sum p=1.7×10−5, CA-Max p=7.9×10−5). Gene-wise CA-Sum and -Max results suggest that the associations are driven by one SNV in each gene; the most significant are common synonymous SNV in TLR6 (T361T and K421K, in strong LD) and ENG (L69L). Although the biological significance of these SNVs remains unclear, their association may provide evidence for rare risk factors for complicated SAB. TLR6 (dimerized with TLR2) recognizes S. aureus cell surface proteins, activating innate immune responses. Mutations in ENG, a cell surface receptor involved in TGF-beta signaling, cause hereditary hemorrhagic telangiectasia, in which SAB is a common comorbidity. A cell surface receptor involved in TGF-beta signaling, cause S. aureus cell surface proteins, activating innate immune responses. Mutations in ENG, a cell surface receptor involved in TGF-beta signaling, cause hereditary hemorrhagic telangiectasia, in which SAB is a common comorbidity.

1144F Implication of CD1H1 in two familial cases of nonsyndromic cleft lip with or without cleft palate. L.A. Brito, G. Yamamoto, M. Aguena, M.R. Passos-Bueno. Department of Genetics, Institute of Biosciences, University of São Paulo, São Paulo, Brazil.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a complex trait and the most common craniofacial birth defect. Familial cases account for 20-30% of the cases, mostly presenting a non-generative pattern of inheritance. Association of common variants has been reported, but they are far from explaining the whole heritability of the disease. In this sense, rare variants might help to bridge the missing heritability gap. We carried out the genome sequencing of 4 affected individuals (HisScan SO - Illumina Inc; mean coverage of 100x) of a large Brazilian family segregating NSCL/P in an autosomal dominant and incompletely penetrant fashion (family 1), and identified a nonsense mutation in the exon 6 (c.G760A:p.D254N) of E-cadherin gene (CDH1), which had never been described in any population. This mutation is located in a conserved region (LjB-PhyloP observation score = 0.99) and is predicted to damage the protein function (LjB-SIFT score = 1; LjB-Polyphen score = 0.99 [scores ranging from 0 to 1, where 1 is the most deleterious prediction]). We screened, by Sanger method, 222 unrelated controls and 283 unrelated patients, and found the same mutation in only one patient. This patient belongs to a family with multiple affected members, from a region geographically close to that of family 1, raising the possibility of founder effect. Considering both families, we estimated the penetrance at 53%. CDH1 is largely known to be involved with familial diffuse gastric cancer, and it has also been implicated in breast, ovarian, endometrial and prostate cancer. A few authors have reported the presence of cleft lip and/or palate in individuals belonging to families segregating familial diffuse gastric cancer, suggesting a partial common etiology between these traits. Our families do not report any case of gastric cancer, which, in turn, suggests that CDH1 might also be implicated with NSCL/P. Recently, one group described potentially deleterious mutations in CDH1 in NSCL/P individuals from European descent without known history of gastric cancer in their families. In conclusion, we describe the impact of a rare mutation underlying two familial cases of NSCL/P, in a gene which had never been implicated by common-variant association studies. In addition, our results suggest the involvement of CDH1 in NSCL/P and underscore the need for genome sequencing and_association studies approach to better understand the genetic architecture of NSCL/P.

1146W Rare variants and risk for asthma in 7,224 individuals from ethnically diverse populations. C. Igartua1,2, R. Myers1, C. Ober1, D.L. Nicolea1,2,3, The Eve Consortium. 1) Human Genetics, University of Washington, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL.

Asthma is a chronic respiratory disease characterized by inflammation and obstruction of the airways, with heritability estimates between 20-80%. Although many common risk alleles have been identified for asthma, together they explain very little of the heritability. Because the genetic architecture of asthma likely involves a combination of common and rare alleles, we evaluated the role of rare variants on asthma risk in subjects of European, African, and Latino ancestry that are part of the EVE Consortium, a multi-center sample that includes 5,523 case-control and 567 parent-child trios. Subjects were genotyped using the Illumina Infinium Human-Exome Chip array, which contains ~250k variants enriched for rare coding variants that are predicted to affect protein function. A gene-based association test for each gene on the exome chip that had at least 2 functional SNPs (nonsynonymous, stop or splicing) was performed using the optimized SNP-set Kernel Association Test (SKAT-O), and applying the first 2 principal components to adjust for population stratification. We considered a total of 142,744 variants in 13,727 genes in each study sample. Only the PGLYRP4 gene (10 variants) in the African Caribbean sample was significant after multiple testing correction (p=2.04×10−6). When functional variants were weighted based on their allele frequencies (rare variants weighted more), 3 genes approached genome-wide significance in African Americans: RAPGEF6 (6 variants; p=1.1×10−5), CULT7 (19 variants; p=1.2×10−5) and SLC22A5 (12 variants; p=4.9×10−5). We also investigated genes near reported asthma GWAS SNPs and found only one instance (SLC22A5) where the association was driven by exonic variation. No genes were significant in the other ethnic groups or in the combined sample. These results indicate that rare variants may contribute more significantly to asthma in individuals of African ancestry, possibly due to the larger number of rare alleles in those populations. Our study, which was limited to variants present on the Human Exome array and conducted in relatively small samples, does not provide evidence for large-effect functional low-frequency exonic variants in the etiology of asthma.
1147T
Targeted next generation sequencing and functional genomics in alopecia areata identifies ULBP6 as a critical node in its genetic architecture. L. Petukhova 1,2, E. Dril 1,2, Z. Dai 1, L. Bian 1, M. Duvois 1, M. Hordinsky 1, D. Norris 1, V. Price 1, R. Cyns 1, A. M. Christiansen 1,6, L. Norris 1, V. Price 1, R. Cyns 1, A. M. Christiansen 1,6, 1) Dept Dermatology, Columbia Univ, New York, NY; 2) Dept Epidemiology, Columbia Univ, New York, NY; 3) Dept Biostatistics, Columbia Univ, New York, NY; 4) Dept Medicine, Columbia Univ, New York, NY; 5) Dept Genetics & Development, Columbia Univ, New York, NY; 6) Dept Dermatology, MD Anderson Cancer Center, Houston, TX; 7) Dept Dermatology, Univ of Minnesota, Minneapolis, MN; 8) Dept Dermatology, Univ of Colorado, Denver, CO; 9) Dept Dermatology, UCSF, San Francisco, CA.

Alopecia areata (AA) is a highly prevalent and poorly understood autoimmune disease which targets the hair follicle causing disfiguring hair loss. It is an enormous unmet medical need for the 5.3 million patients in the US who suffer from AA, arising primarily from a lack of understanding of disease pathogenesis. Our initial GWAS in AA revealed the first disease association to ULBP3/6 genes in any human disease. These genes are ligands for the NKG2D activating receptors for a repertoire of leukocytes. We biologically validated our statistical evidence by showing a marked increase in CD8+ NKG2D+ T cells within the immune infiltrate. These findings, together with the previous demonstration of MICA overexpression in AA hair follicles, placed the NKG2D axis squarely at the center of AA pathogenesis, and invited a functional genomics approach to uncover causal variants predisposing to disease. In order to better understand the genetic variation driving the tagSNP associations identified in our GWAS at the ULBP3 locus, we selected a subset of 124 cases from our GWAS cohort for targeted deep resequencing with RainDance technology, amplifying 72Kb of sequence encompassing the entire region of association. As preliminary analysis of this dataset, we looked at the distribution of rare variants (p<0.01 in EVS and 1000G) across this region. We identified two rare missense variants, one of which is highly overrepresented in our cohort (p=0.005) and is located within ULBP6. Of the 127 rare or novel variants located within intergenic regions we identified 34 that fall within transcription factor binding sites, 7 of which are overrepresented in our cases, which cluster into two regions. One of these regions is a CTCF binding site, which is known to influence chromosome structure providing a mechanism for the regulation of gene expression. We have begun to assess the biological consequences of the ULBP6 protein coding variant, by developing a battery of functional assays aimed specifically at examining the impaired expression traits, genetic effect on cell surface display, receptor binding affinity and killing efficiency. We are additionally developing cellular assays to determine the effects of the regulatory variants. This work will clarify how GWAS identified genetic variation influences NKG2D-mediated cytotoxicity in the pathogenesis of autoimmune disease.

1148F
Exploring the genome for the secrets of human longevity. E.B. van den Akker 1,2, S.J. Pitts 3, M.H. Moed 1, S. Pollan 2, J. Deelen 1,4, J.J. Houwing-Duistermaat 1,2, D.R. Cox 3, M.J.T. Reinders 1, M. Beekman 1,4,7, P.E. Slagboom 1,4,7, Genome of The Netherlands Consortium. 1) Molecular Epidemiology, Leiden University Medical Center, Leiden, Zuid Holland, The Netherlands; 2) The Delft Bioinformatics Lab, Delft University of Technology, Delft, Zuid Holland, The Netherlands; 3) Rina-Pfizer Inc, South San Francisco, CA 94080, USA; 4) Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands; 5) Leiden University Medical Center, Medical Statistics and Bioinformatics, Leiden, The Netherlands; 6) In Memoriam; 7) These authors contributed equally to this work.

The clustering of longevity in families is being studied worldwide and the genetic contribution to longevity in the population at large is estimated at 25%. This is illustrated by the fact that first-degree family members of nonagenarians and centenarians also have a life-long survival advantage that can be attributed to a lower risk of coronary artery disease, cancer and type-2 diabetes. The survival advantage in 420 Dutch long-lived families from the Leiden Longevity Study (LLS), which have been selected on the basis of living nonagenarian siblings (≥2), could not be explained by the absence of risk alleles at known disease susceptibility loci. To investigate whether rare private genetic variants contribute to their longevity, we explored the whole-genome sequence of 220 unrelated Dutch nonagenarians (mean age 94 years) in comparison with 95 younger controls from the general Dutch population. The nonagenarians were selected from the LLS study population on the basis of having the best family history of survival and the controls were randomly selected from Dutch biobanks included in the BBMRI-NL consortium consisting of healthy individuals. High quality whole-genome sequences of DNA from leukocytes were generated by Complete Genomics (USA) at >30x coverage. The mean numbers of SNPs, deletions, insertion and substitutions per individual are remarkably similar between controls and the nonagenarians, which may be regarded as super-controls. The mean numbers of variants observed in the exome of controls were 3,344,320.57 SNPs, 224,143 deletions, 217,134 insertions and 81,475 substitutions and in nonagenarians 3,357,346.38 SNPs, 225,241 deletions, 213,847 insertions and 80,436 substitutions, respectively. Further, we investigated differences between the two groups in the presence of longevity alleles at known loci (a.o. TOMM40/APOE and FOXO3A). Next, we explored all genes in a burden analysis and paid special attention to IIS and mTOR pathway related genes. For this analysis we selected all detected genetic variants with a minor allele frequency (MAF) below 1% according to ANNOVAR. Our preliminary data analysis indicated a tendency that for many exonic annotations the nonagenarians carry fewer variants than controls.
Simulations of Finnish population history, guided by empirical genetic data, to assess power of rare variant tests in Finland. R. Wang1,2,3, A. Agarwala2,4,5, J. Flannick2,3,6, D. Altshuler2,3,6, J.N. Hirschhorn1,2,3, 1Division of Endocrinology, Boston Children’s Hospital; 2Program in Medical and Population Genetics, Broad Institute; 3Department of Genetics, Harvard Medical School; 4Harvard-MIT Division of Health Sciences and Technology; 5Program in Biophysics, Graduate School of Arts and Sciences, Harvard University; 6Department of Molecular Biology, Massachusetts General Hospital.

The Finnish population has been extensively utilized in genetic studies. It is considered to be a relatively homogeneous founder population, which is hence well suited for gene mapping, especially for variants of lower frequency. However, without realistic population demographic and phenotypic models, it is difficult to assess the implications of the founder effect for association studies. Studies of rare variants in exome genotype or sequence data. In this study, we developed a population genetic framework to address these issues. We first confirmed the Finnish founder effect in deep resequencing data by showing that allele frequency spectra are shifted towards higher end in the Finns compared to non-Finnish Europeans (NiEEs). Next, building on recent simulations that generate representative sequence data for European populations, we used empirical sequence data to develop a simultaneous forward simulation of sequence variation in the NFE population and the Finnish population. We then simulated phenotypes in these populations under different disease models characterized by different correlation between phenotypic effects of variants and their effects on fitness. With simulated genotype and phenotype data, we conducted rare variant statistical tests (single variant, burden tests), and assessed their performance by calculating power. Single variant tests showed a 4.47 power boost in Finns compared to NiEEs. Finns could provide additional power especially when phenotypic effects of variants correlate well with effects on fitness. The single variant test shows good performance in a founder population under different disease models, and it is particularly powerful if rare variants play a big role in explaining phenotypic variance. Finally, we compared the efficiency of genotype data (exome chip) and sequence data (exome sequencing) under different scenarios. Our results suggest that exome chip is currently much more cost-efficient than exome sequencing, especially in a founder population. Our study has highlighted the usefulness of understanding the population-genetic properties of a study population and exploring a range of genetic models for recognizing the features and limitations of association studies in that population. Further work is needed to focus on rare variant discovery and expanded sources of variation, such models offer a context with which to interpret the data and to plan future studies for gene discovery.

1150T
De novo and inherited retrotransposon insertions associated with autism revealed by whole genome sequencing. X. Jin1,2, A. Ewing1, J. Ju1, R. Yuen1, J. Wu1, Y. Jiang1, M. Wang1, A. Shih2, Y. Li1, J. Wang1, S. Scherer2, Z. Sun1, H. Yang1, H. Kazazian1, J. Wang1, BGI, Shenzhen, Guangdong, China; 2School of Biomedical Sciences, University of Southampton, Southampton, United Kingdom.

Retrotransposons are genetic elements that are capable of moving by a ‘copy-and-paste’ mechanism to spread throughout genome. Around 45% of the human genome is composed of retrotransposons. Recent studies have revealed that somatic mosaicism driven by retrotransposon insertions can reshape the genetic circuitry that underpins normal and abnormal neurobiological processes. Autism spectrum disorder (ASD) is a lifelong developmental condition that affects about 1 in 100 individuals. Multiple researches have confirmed the contribution of de novo and rare inherited mutation to the risk for ASD. Here we sequenced the whole genome of 232 individuals from 74 families with ASD collected in China and the US. We developed a novel method to identify de novo and inherited retrotransposon insertions from paired-end whole-genome sequencing data. We identified 2 high-confidence de novo retrotransposon insertions in genes MARK2 and RIMS2, both of which have been reported to contain de novo functional mutations in recent large-scale exome studies of autism. We also detected 1281 to 1725 (average 1530) inherited retrotransposon insertions in each family. Three of these de novo or inherited retrotransposon insertions were located in known autism related genes such as CTNNA3, AUTS2 and TCF4. Our findings suggest that de novo and inherited retrotransposon insertions may play an important role in autism genetics.

1151F
Exome sequencing of 55 multiply affected coeliac families and large scale resequencing follow up. V. Mistri1, A. N.A. Bockett1, M. Muddasar2, K.A. Hunt3, S.L. Nevanlinna4, P.J. Ciclitira3, V. Plagnol4, D.A. van Heel4, 1Centre for Digestive Diseases, Barts and The London School Of Medicine and Dentistry, Bizard Institute, 4 Newark Street, Whitechapel, London E1 2AT, United Kingdom; 2Division of Genetics and Molecular Medicine, Kings College London School of Medicine, 8th Floor Tower Wing, Guy’s Hospital, London SE1 9RY, United Kingdom; 3Department of Population Sciences, City of Hope, Duarte, California 91010, USA; 4King’s College London, Division of Diabetes and Nutritional Sciences, Gastroenterology, The Royal Inns of Chelsea Hospital, Fulham Road, London, United Kingdom; 5University College London Genetics Institute, Gower Street, London WC1E 6BT, United Kingdom.

Coeliac disease is a highly heritable common autoimmune disease involving small intestinal inflammation in response to dietary wheat. The HLA region, and 40 newer regions identified by GWAS and dense fine mapping (many immune genes), account for ~40% of heritability. We hypothesized that rare mutations of larger effect size (~OR ~ 2 - 5) might exist, especially in multiply affected pedigrees. We exome sequenced 75 subjects from 55 multiply affected families. We selected interesting variants/genes for further follow up using a combination of: linkage, shared variants between multiple related subjects and gene burden tests for multiple potentially causal variants. We next performed highly multiplexed amplicon sequencing (Fluidigm) of all RefSeq exons from 24 candidate genes in 2,304 coeliac cases and 2,304 controls. High coverage data enabled direct genotyping (99.98% of all sample genotype calls had a read depth >40) and extensive quality control. 1,335 unique variants with a 99.98% genotyping call rate were called (3-4x fold of previous studies), 1,472 of which were present in coding regions of 24 genes (TvTv 2.99). 91.7% of coding variants were rare (MAF in 2,350 controls, <0.5%) and 60% were novel. No common or low frequency variants were seen at novel sites. Gene burden tests (C-Alpha/LAT) performed with rare functional variants identified no rare significant associations (P<1x10-5) at the resequenced candidate genes. Our strategy of sequencing multiply affected families, and deep follow up of candidate genes, has not identified new disease risk mutations. Common variants and other factors, e.g. environmental, may instead account for familial clustering in this common autoimmune disease.

1152W
Pathway analysis using whole exome sequencing in Parkinson disease. K. Nytenm2, V. Inchausti1, L. Maldonado1, W. Perry1, E.R. Martin1,2, G.W. Beecham1,2, L. Wang1,2, W.K. Scott1,2, J.M. Vance1,2, 1John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA.

Previously, we reported several biological pathways (KEGG) to be associated with Parkinson Disease (PD) using gene expression and GWAS data (Edwards 2011). The top 3 pathways included ‘axon guidance’, ‘focal adhesion’ and ‘calcium signaling pathway’. These analyses were performed on 196 PD cases and 234 controls. Results suggested that common, often intergenic or intronic genomic variants with low risk effects and downstream differences in transcript level. In contrast, variants identified in whole exome sequencing were enriched for ‘rare risk variants’, variants with a low frequency but possible higher risk effect on disease. We set out to determine whether the combination of exome variants in the same pathways also provides evidence for association with increased PD risk. We performed WES in 315 PD patients and 344 controls (all unrelated). We identified 2 sets of variants in pathways with increased risk for PD was assessed using the Cochran-Armitage sum and max test implemented in the RVASSOC program and the Sequence Kernel Association test (SKAT). Analyzed pathways included 13 top KEGG pathways identifying the previous reported pathway study as well as candidate gene groups ‘axon transport’, ‘mitochondrial’ (Mitocarta) and ‘lysosomal’ genes (Human Lysosome Gene Database). On average, ~85% and ~73% of exome variants in these pathways have a frequency <5% and <1% respectively. Using minimal significant results (0.05<p<0.01) were obtained for 5 out of 13 top reported KEGG pathways in PD. However, we observed evidence for association with PD risk for the ‘mitochondrial’ and ‘lysosomal’ gene groups (p<0.001). Permutation tests (N=500) on random sets of the same number of genes or variants as the original groups identified that the original group was significantly more associated than the random gene sets. Interestingly, when filtering on variant frequency (<5% or <1%) we obtained similar results for all but the ‘lysosomal’ gene group. A significant p-value for this group was observed in both rare variant analysis (p-value<0.01) and >5% analysis (p-value<0.02). Our findings were further validated by our previously reported pathways with PD risk could be identified when analyzing all exome variants. Further filtering on frequency indicated that rare variants in the ‘lysosomal’ gene group are significantly contributing to disease. Additional analyses on smaller candidate group with variants filtered on function will be performed to elucidate the observed signal in this analysis.
1153T Whole-genome sequencing in a Multiplex Nuclear Family of Schizophrenia to Identify Its Rare Susceptibility Variants, S.C. Yu1, H.Y. Chen2, S.L. Yu3, W.H. Chen4, J.T. Ho4, G. Wu5, R.H. Hirota2, T. Sakurai1, Y. Takanashi1, A. Reijo Pahikkala1,5, K. Takanashi1, Y. Hori2, F. Fujita3, Y. Suzuki1, Y. Fujieda1, N. Nagano5, K. Ichimura1, M. Hayakawa1, Y. Inazawa1, A. Gollapudi6, S. Iwata1, J. Gruis6, R. Feenstra6, A. Luengo6, B.C. St Clair7, N.G. Hindorff7, J. Bevreg8, S. Andreassen8, K. Schork10, K. Watanabe11, W. Hoek12,13,14, P. van den Borne15, J. de Geus11, S. Sullivan16, P. Table12, H. Kojima1,4, J. van Duijn17, D. Torgerson18, M. Maher19, D. E. Martin20,21, A. Palotie22,23. 1) Institute of Epidemiology and Preventive Medicine, College of Public Health, Nat, Taipei, Taiwan; 2) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; 1315, ROC; 3) Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University College of Medicine, 1 Chang- te Street, Taipei, 100, Taiwan; 4) Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University; 5) Graduate Institute of Brain and Mind, College of Medicine, National Taiwan University; 6) National Taiwan University; 7) Department of Medicine, University of California, San Francisco, CA; 8) Department of Neurology, University of Pennsylvania, Philadelphia, PA; 9) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 2) Department of Pediatrics, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 3) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

Background: Genome-wide association study (GWAS) has limitations for exploring rare genetic variants. With the advent of next-generation sequencing (NGS), it becomes possible to search for the whole genome for rare variants specific to schizophrenia patients. A previous study among a large number of families of sib-pair co-affected with schizophrenia in Taiwan has revealed several linkage signals by incorporating several endophenotypes. This implies existence of genetic heterogeneity in schizophrenia. We hence postulated that high density nuclear families of schizophrenia may help identify certain inherited rare variants that can be used for replication in other multiplex families of schizophrenia. Method: We selected one high density schizophrenia family from Taiwan Schizophrenia Linkage Study (TSLS), which recruited schizophrenia patients and their first-degree relatives throughout Taiwan from 1998 to 2002. NGS was performed to sequence the whole genomes of a 5-member nuclear family, in which the mother and 2 children affected with schizophrenia, and the father and another unaffected daughter. Only 7 variants were observed in both dominantly and recessively inherited samples were used to explore schizophrenia related genomic variations. Results: The results showed that total 384 variants (2 exonic single nucleotide variants; SNVs) selected under recessive inheritance model, i.e., with schizophrenia, has a higher frequency of rare variants and both recessive and heterozygous variants. In dominant inheritance model, there were 1371 variants (70 exonic SNVs) selected, i.e., with schizophrenia patients having the variant and health persons having wild-type allele only. In addition, 10 non-synonymous exonic SNVs that exhibited inheritance in the family but not seen in 1000 genomes project were identified for those in accordance with dominant inheritance model. Discussion: Our findings demonstrated the utility of NGS in identifying inherited rare genetic variants that are potentially associated with multiplex schizophrenia. Conclusion: We may finally identify the key genetic factors driving complex diseases.

1155W Targeted sequencing of the pericentromeric region of chromosome 2 in Finnish constitutional delay of growth and puberty families. D. Courteille1, L. Dunker2, A. Palotie3,4, E. Widén4, 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 2) Department of Pediatrics, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 3) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

Constitutional delay of growth and puberty (CDGP) is the most common cause of pubertal delay, representing about 65% of boys and 30% of girls with late puberty. CDGP is defined as the absence of pubertal development at an age 2 standard deviations above the population mean of normal pubertal timing. While both environmental and genetic factors contribute to variability of pubertal timing and tempo, high heritability estimates of up to 80% predict a strong influence from genetic factors. At the tail ends of the population distribution, one or several genes with large effect sizes are predicted to have a strong influence on the phenotype.

Recent studies showed that the majority of Finnish families segregated CDGP in an autosomal dominant manner, and a subsequent linkage study identified a susceptibility locus in the pericentromeric region of chromosome 2. We followed up on this linkage signal by performing targeted sequencing of over 60 Mb under the linkage peak in the proband and affected parent of the 13 best-linked families. Poorly sequenced regions were imputed against Finnish 1000 Genomes reference samples. Following sequencing and imputation, transmitted coding variants and non-coding regulatory variants will be compared among the probands to create a list of potentially causative candidate variants. These candidate variants will be assessed for association with pubertal delay by genotyping independent cases and controls. Results: The first three families of multiplex families were genotyped, demonstrating that the probands are 100% homozygous for the same variants. This may lead to discovery if rare but inheritable susceptibility variants for CDGP. Our use of a multiplex family setup might have helped exclude those variants due to typing error, or de novo mutations, and hence select those with high-penetrance susceptibility genetic variants for schizophrenia. In particular, the 10 non-synonymous exonic SNVs warrant future replication in other multiplex families from TSLS. This may lead to discovery if rare but inheritable susceptibility variants for schizophrenia and their underlying pathophysiology.

1156T Population genetics of rare variants and complex diseases. R. Hernandez1, M. Maher2, L. Uricchio3, D. Torgerson2, 1) Bioeng. & Therapeutic Sci, UCSF, San Francisco, CA; 2) Department of Medicine, UCSF, San Francisco, CA.

Identifying drivers of complex traits from the noisy signals of genetic variation has proved challenging from high throughput genome sequencing technologies. Population genetic differentiation among ethnic groups is a central challenge faced by human geneticists today. We hypothesize that the variants involved in complex diseases are likely to exhibit non-neutral evolutionary signatures. Uncovering the evolutionary history of all variants is therefore of intrinsic interest for complex disease research. However, doing so necessitates the simultaneous elucidation of the targets of natural selection and population-specific demographic history. Here we characterize the action of natural selection operating across complex disease categories, and use population genetic simulations to evaluate the expected patterns of genetic variation in large samples (n=10,000). We focus on populations that have experienced historical bottlenecks followed by explosive growth (consistent with many human populations), and describe the differences between variants associated with diseases occurring in European Americans versus individuals of African descent. We find that the genes associated with several complex disease categories exhibit stronger signatures of purifying selection than non-disease genes. In addition, loci identified through genome-wide association studies of complex traits also exhibit signatures consistent with being in regions recurrently targeted by purifying selection. Through simulations, we show that population bottlenecks and rapid growth enables deleterious rare variants to persist at low frequencies just as long as neutral variants, but low frequency and common variants tend to be much younger than neutral variants. This has resulted in a two-stage model for rare variants that have a deleterious effect on function, and that potentially contribute to disease susceptibility. The key question for sequencing-based association studies of complex traits is how to distinguish between deleterious and benign genetic variation. We show how Bayesian modeling and a suite of tests that distinguish these two categories, especially derived allele age, thereby providing inroads into novel methods for characterizing rare genetic variation driving complex diseases.
1157F
Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders.

H.N. Cuiker1, N.D. Duerker1, S.H. Silfver1, J.M. Lee1, P.L. Whitehead2, E. Lalanne3, N. Leyva4, I. Konidari5, R.C. Gentry6, W.F. Huyme1, D. Van Booven1, V. Mayo1, N.K. Hofmann2, M.A. Schmidt2,2, E.R. Martin2,2, J.L. Haines2, M.L. Cuccaro1,2, J.R. Gilbert1,2, M.A. Pencak-Vance1,2,1, 1) John P. Hussman Institute for Human Genetics, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA.

Autism spectrum disorders (ASDs) encompass a constellation of neurodevelopmental conditions and studies to date demonstrate that the underlying etiology is extremely heterogeneous. To help unravel this genetic complexity, we performed whole exome sequencing in 40 ASD families with multiple, distantly related, ASD affected individuals. In contrast to previous autism exome studies which have primarily focused either on simplex families to discover de novo alterations or consanguineous families that carry recessive mutations, our study required that each family contain at minimum one pair of affected cousins. A total of 164 individuals were captured with the Agilent SureSelect Human All Exon kit, sequenced on the Illumina HiSeq 2000, and the resulting data processed and annotated with BWA, GATK, and SeattleSeq. Each family had approximately 90,000 changes. Variants were filtered to those in identity by descent (IBD) regions delineated by SNP genotyping data. Initial analyses focused on novel and rare (MAF < 0.05) variants predicted to be detrimental, either by altering amino acids or splicing patterns. In accordance with a dominant model of inheritance, exome sequencing identified 742 heterozygous changes, 499 of which were validated either by Sanger sequencing or genotyping on the Illumina HiSeq 2000, and the resulting data processed and annotated with BWA, GATK, and SeattleSeq. Each family had approximately 90,000 changes. Variants were filtered to those in identity by descent (IBD) regions delineated by SNP genotyping data. Initial analyses focused on novel and rare (MAF < 0.05) variants predicted to be detrimental, either by altering amino acids or splicing patterns. In accordance with a dominant model of inheritance, exome sequencing identified 742 heterozygous changes, 499 of which were validated either by Sanger sequencing or genotyping on the Illuminum HumanExome BeadChip. We identified numerous potentially damaging, ASD associated risk variants in genes previously unrelated to autism. A subset of these genes has been implicated in other neurological disorders including depression (SLIT3), epilepsy (CLCN2, PRICKLE1), mental retardation (AP4M1, CEP290), schizophrenia (WDR60), and Tourette syndrome (OFCC1). This reinforces the theory that there are shared genetic components across distinct neurological disorders. Additional alterations were found in genes with significant overrepresentation, including three genes with alterations in multiple families (CSMD1, FAT1, and STXBP5). Compiling a list of ASD candidate genes from the literature, we determined that variants occurred in ASD candidate genes 1.65 times more frequently than in random genes captured in this study (p = 2.0 × 10−3). By studying these unique pedigrees, we have identified novel DNA variations related to ASDs, demonstrated that exome sequencing in extended families is a powerful tool for ASD candidate gene discovery, and provided further evidence of an underlying genetic component to a wide range of neurodevelopmental and neuropsychiatric diseases.

1158W
Nextgen RNA sequencing of monocytes coupled with association data identifies several genes in Systemic Lupus Erythematosus susceptibility.

A. Maiti1, J.K. Grange2,3, R.A. Yang4, C. Buchan5, D. Maiti1, A. Alvarado5, P. Yang4, C. Crucigma4, M.B. Harms4, K. Ha5, T. Zhang1, M.C. Willing5, D.K. Grange5, T.E. Druley1,2,6, A.C. Braverman7, M. Lovett1, M.B. Dobbs2,7, C.A. Gurnett2,3,5, 1) Genetics, Washington University, Saint Louis, MO; 2) Neurology, Washington University, Saint Louis, MO; 3) Orthopedic Surgery, Washington University, Saint Louis, MO; 4) Psychiatry, Washington University, Saint Louis, MO; 5) Pediatrics, Washington University, Saint Louis, MO; 6) Medicine, Washington University, Saint Louis, MO; 7) St. Louis Shriners Hospital for Children.

Adolescent idiopathic scoliosis (AIS) is a common pediatric spine deformity that affects up to 3% of the population. Despite a strong genetic component, few genes have been associated with AIS and the pathogenesis remains poorly understood. Marfan syndrome and congenital contractual arachnodyasty are Mendelian disorders that are highly associated with scoliosis and are caused by mutations in the large fibrillin genes, FBN1 and FBN2. To determine whether rare genetic variants in fibrillin contribute to the pathogenesis of AIS, we sequenced FBN1 and FBN2 using a novel and cost-effective method of targeted capture called multiplex direct genomic sequencing (MDGS) in combination with exome sequencing. In individuals of European ancestry, an increased frequency of private FBN1 and FBN2 variants was identified in AIS cases (8.2%; n=233) compared to controls (2.0%; n=393) (OR=4.27; p=2.7×10−4). Private FBN1 and FBN2 variants were associated with more severe spinal deformity (p=3.2×10−4), but were not associated with systemic features of Marfan syndrome or congenital contractual arachnodactyly. Activation of the TGF-β signaling pathway was demonstrated by elevated pSMAD2 in muscle of AIS patients with private FBN1 and FBN2 variants. Our findings underscore the functional importance of rare fibrillin variants in AIS pathogenesis and provide the possibility of novel therapeutic strategies for treating AIS.

1159T
Multiplexed targeted capture of FBN1 and FBN2 reveals association with adolescent idiopathic scoliosis.

J.G. Buchan1, D.M. Alvarado2, P. Yang2, C. Crucigma4, M.B. Harms4, K. Ha5, T. Zhang1, M.C. Willing5, D.K. Grange5, T.E. Druley1,2,6, A.C. Braverman7, M. Lovett1, M.B. Dobbs2,7, C.A. Gurnett2,3,5, 1) Genetics, Washington University, Saint Louis, MO; 2) Neurology, Washington University, Saint Louis, MO; 3) Orthopedic Surgery, Washington University, Saint Louis, MO; 4) Psychiatry, Washington University, Saint Louis, MO; 5) Pediatrics, Washington University, Saint Louis, MO; 6) Medicine, Washington University, Saint Louis, MO; 7) St. Louis Shriners Hospital for Children.

Adolescent idiopathic scoliosis (AIS) is a common pediatric spine deformity that affects up to 3% of the population. Despite a strong genetic component, few genes have been associated with AIS and the pathogenesis remains poorly understood. Marfan syndrome and congenital contractual arachnodyasty are Mendelian disorders that are highly associated with scoliosis and are caused by mutations in the large fibrillin genes, FBN1 and FBN2. To determine whether rare genetic variants in fibrillin contribute to the pathogenesis of AIS, we sequenced FBN1 and FBN2 using a novel and cost-effective method of targeted capture called multiplex direct genomic sequencing (MDGS) in combination with exome sequencing. In individuals of European ancestry, an increased frequency of private FBN1 and FBN2 variants was identified in AIS cases (8.2%; n=233) compared to controls (2.0%; n=393) (OR=4.27; p=2.7×10−4). Private FBN1 and FBN2 variants were associated with more severe spinal deformity (p=3.2×10−4), but were not associated with systemic features of Marfan syndrome or congenital contractual arachnodactyly. Activation of the TGF-β signaling pathway was demonstrated by elevated pSMAD2 in muscle of AIS patients with private FBN1 and FBN2 variants. Our findings underscore the functional importance of rare fibrillin variants in AIS pathogenesis and provide the possibility of novel therapeutic strategies for treating AIS.
Novel candidate genes putatively involved in stress fracture predisposition detected by whole Exome sequencing. E. Friedman1,2, D.S. Moran1, D. Ben-Avraham1, R. Yanovich2, G. Atzmon3, 1) Oncogenetics Unit, Inst Gen, Chaim Sheba Med Ctr, Tel Hashomer, Israel; 2) The Sackler School of Medicine, Tel-Aviv University, Ramat Aviv; 3) The Military Physiology Unit, Heller Institute, Sheba Medical center, Israel; 4) Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY 10463.

While clearly genetic factors are involved in stress fracture (SF) pathogenesis, few studies that focused on candidate genes reported on the contribution of sequence variants in these genes to this common overuse injury. We employed an unbiased screening approach by using exome sequence capture arrays followed by next generation sequencing (HiSeq2000) of two pooled DNA samples from Israeli soldiers: cases with high grade SF (n = 34) and ethnically and age matched controls with no evidence of SF (n = 60). Of the 144,217 and 202,406 sequence variants (control and case pools, respectively) 67,408 variants passed the various QC filtering stages with 3-20,000 reads/variant. Of 1900 variants with more than 600 reads/variant in both DNA pools, 145 sequence variants (in 127 genes) displayed statistically significant (p<0.05) differences in rates between cases and controls. Subsequent validation of these 145 sequence variants individually for the 55 controls and 32 SF soldiers who formed the pooled samples using the Sequenom platform, validated 11/145 SNPs. A second, independent, individually genotyped validation cohort with 72 controls and 104 SF soldiers using the Sequenom platform to query the same 145 SNPs, validated only 1 SNP: a missense mutation in the BTN3A1 gene. Combined analysis of the two datasets for all cases and controls with adjustment for inter-batch variability resulted in 118 SNPs. Secularly distant cases and controls, 5 of which were in the originally detected and validated 11 SNPs. Of these, three SNPs were Synonymous SNPs located within the SEC24D, 1 SNP in the LOM1 gene. In conclusion, exome sequencing of DNA pools provided novel candidate genes seemingly involved in SF pathogenesis and predisposition.

A possible role of Transposable Elements in dysregulating the genomic architecture of Schizophrenia. F. Macciardi1,2, G. Guffanti3, S.G. Potkin4, M. Pato5, I. Guella1, M. Vawter6, A. Knowles7, T. VanErp8, C. Pato2, S. Gaud1. 1) Dept. of Psychiatry & Human Behavior, University of California Irvine (UCI), Irvine, CA; 2) Department of Psychiatry Division of Epidemiology & Division of Child and Adolescent Psychiatry Columbia University/NYSPH, New York, NY; 3) Department of Psychiatry and the Behavioral Sciences - Keck School of Medicine at USC, Los Angeles, CA; 4) Department of Infectious, Parasitic and Immune-Mediated Diseases, Italian National Institute of Health, Rome, Italy.

Background. Between half and two-thirds of our genome is composed of repetitive elements, the complexity Transposable Elements (TEs, e.g, LINEs, SINES, SVAs and HERVs). Since their discovery TEs were thought to act as ‘controlling elements’ of nearby genes. While the notion that DNA could be mobile was accepted, the idea of control was not. The role of TEs as regulatory elements is now more recognized. TEs create variability by retro-transposition Insertion Polymorphisms and by SNVs in fixed TEs. By retrotransposing to new insertional sites, TEs create structural variants and provide novel promoters, splice sites, exons or polyadenylation signals. Methods. We identified and annotated SNVs and Retrotransposition Insertion Polymorphisms (RIPs) for Insertion Polymorphisms from various classes of TEs (LINEs, SINE, HERVs and SVAs) from whole-genome sequence data in post-mortem brains of 10 cases and 10 controls. We examined TE sequence differences (including RIPs) between SZ patients and controls, using re-alignment and de novo assembly and characterized the genomic context of RIPs (exon, introns, 5’ and 3’ UTRs, non-coding regions). Results. We found that RIPs are largely underrepresented in SZ patients compared to controls for all TE classes. The number of RIPs in controls ranges from 3,630 to at least 4,500, within the boundaries calculated from the 1,000 Genome Project for dbRIP-reference and RIP non-reference TEs, a larger estimate than from previous studies. In cases RIPs are 1/10 to 1/20 than in controls. Discussion. The low number of RIPs in SZ subjects suggests a neuro-developmental defect affecting embryonic brain development. Recent stem cell models show that a high number of RIPs for L1s and Alus is a key factor for somatic neuronal development, although the precise mechanisms are not yet clear. It is thus possible that a reduction of RIPs in SZ at the early embryonic level may be less relevant than the somatic retrotransposition at later development that occurs in neurons at different stages and results in an increased transcriptome heterogeneity among neurons. Our preliminary results reveal a new, unexpected dimension of the regulatory genome that may play a pivotal role in the etiology of schizophrenia. Since their identification, TEs represent a controversial concept due to their biology and mechanisms of action that challenge genetic dogma. Our findings, however, suggest that TEs can represent an important new genomic risk factor in schizophrenia.

ImmunoSeq: Discovery of novel rare variants implicated in autoimmune and inflammatory diseases by targeting regulatory regions in immune cells. A. Morin1,2, T. Kwan1,2, K. Tandrea3, M.L Eloranta1, V. Arsenault4, M. Caron1,2, L. Létourneau2, C. Wang2, G. Bourque1,2, C. Laprise5, A. Montpetit2, A.C. Svyänen5, L. Rönnblom5, M. Lathrop5, T. Pasinlenn5, 1) Department of Human Genetics, McGill University, Montréal, Québec, Canada; 2) McGill University and Genome Québec Innovation Centre, Montréal, Québec, Canada; 3) Department of Medical Sciences, Section of Rheumatology, Uppsala University, Uppsala, Sweden; 4) Département des sciences fondamentales, Université du Québec à Chicoutimi Saguenay, Québec, Canada; 5) Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

Genome-wide association studies (GWAS) have identified many common SNPs associated to complex traits, yet they only partially explain their genetic component. Whole-exome sequencing has been used in the discovery of rare variants associated to complex traits, but its potential success is limited, since GWAS have shown that the associated SNPs are predominantly located in noncoding regions. A recent study showed that rare coding variants have limited impact on the development of autoimmune diseases. Comprehensive DNase I hypersensitive site (DHS) mapping by the ENCODE project identified all classes of cis-regulatory elements, and recent studies showed enrichment of GWAS noncoding variants in DHS. The discovery of rare variants located in regulatory regions specific to immune cells should explain part of the heritability of autoimmune complex traits. Our discovery of rare variants located in noncoding regions. A recent study showed that rare coding variants have limited impact on the development of autoimmune diseases.
1163F

AMD risk and association with variation in 202 drug target genes. P.L. St Jean1, W.H. Cade2, F. Grassmann2, M. Schu4, S. Silfer2, Z. Ye2, M.H. Brilliant3, M.M. DeAngelis4, L.A. Farner1, J.L. Haines2, T.E. Kittner1, M.A. Pericak-Vance2, B.H.F. Weber3, L. McCarthy1, C.-F. Xu5, M.G. Ehm1. 1) Quantitative Sciences, GlaxoSmithKline, RTP NC, USA; Stevenage UK; 2) University of Miami Miller School of Medicine, Miami, FL, USA; 3) Institute of Human Genetics, University of Regensburg, Regensburg, Germany; 4) Boston University Schools of Medicine and Public Health, Boston, MA, USA; 5) Marshfield Clinic Research Foundation, Marshfield, WI, USA; 6) Department of Ophthalmology and Visual Sciences, Moran Eye Center, University of Utah, Salt Lake City, UT USA; 7) Center for Human Genetics Research, Vanderbilt University Medical School, Nashville, TN, USA.

Age-related macular degeneration (AMD) is a common cause of blindness and visual impairment in older adults occurring in ~12% of people over age 80. Advanced forms of AMD are geographic atrophy (GA) involving retinal pigment epithelial atrophy, and neovascular (NV) complications characterized by abnormal blood vessel growth and leakage of blood and protein. Approved treatments of AMD, focused on the neovascular forms, include ranibizumab and aflibercept which are both administered by injection. Developing drugs effectively treating GA, as well as non-injectable treatments for neovascular forms, would have a substantial, positive impact for people with advanced AMD. Our objective is to identify genetic variants in drug targets gene that are associated with AMD risk with the hope that some of these genes may prove to be effective targets for developing therapeutics for AMD. Exons and flanking regions of 202 genes encoding current or prospective GSK drug targets were sequenced in 14,000 European subjects. Illumina 500k or Affymetrix 500k data were available on 6900 of these subjects and several AMD collections of European origin. Association between AMD risk and dosage was conducted on a total of 2,696 cases and 3,335 controls for variants with an R2 quality metric threshold of ≥0.5. A meta-analysis of effect estimates was performed using the inverse variance method in METAL while testing for heterogeneity across studies. A predefined statistical threshold of p=3.65E-5, accounting for the effective number of variants, was used to identify significant results while a less stringent threshold of p<0.005 was used to identify suggestive results across the collected, 46 bp IVS variants with an R2 quality of 0.5 were analyzed. Thirteen variants were associated at p=3.65E-5 and these reside in regions flanking, but not within the 202 target genes. Ten of these variants are on chromosome 6 and map to reported AMD risk genes C2, CFB, SKIV2L; however, they are 400-800 kb from the closest GSK target gene. Several variants within GSK target genes were associated at p<0.005 and many of these genes are implicated in AMD disease mechanisms such as inflammation and lipo-protein processing. Further support of the involvement of these target genes in AMD risk will be assessed in independent replication studies.

1164W

Matrix metalloproteinase 2, 3, 9, 10, 13 gene polymorphisms and risk for polycythemia and essential thrombocytosis patients. E. Uctepe1, M.M. DeAngelis2, O. Bender1, T. Yasar1, E. Gunduz1, M. Gunduz1,2, J.L. Haines3, L.A. Farner1, J.L. Haines2, T.E. Kittner1, M.A. Pericak-Vance2, B.H.F. Weber3, L. McCarthy1, C.-F. Xu5, M.G. Ehm1. 1) Medical Genetics Department, Turgut Ozal University, Ankara, Turkey; 2) Internal Medicine Department, Fatih University, Ankara, Turkey; 3) Departments of Otolaryngology Head and Neck Surgery, Turgut Ozal University, Ankara, Turkey.

Chronic myeloid disorders such as polycythemia vera (PV), essential thrombocytosis (ET) and idiopathic myelofibrosis arises from clonal proliferation of neoplastic stem cells in the bone marrow. Matrix metalloproteinases (MMPs) are a family of zinc-dependent endopeptidases that have potential to degrade all types of extracellular matrix (ECM) and also play a role in remodeling of the ECM. It is known that MMPs play a role in bone marrow fibrosis. The primary goal of our study is to determine the relationship between chronic myeloproliferative diseases and MMP gene polymorphisms. The demonstration of a relationship will determine whether polymorphisms lead to susceptibility to disease and with future work may aid in the development of new therapeutic modalities. Patients were selected from outpatient clinics of the Turgut Ozal Medical University Hospital between December 2010 and May 2011. Twenty-eight patients who previously diagnosed and follow-up with a diagnosis of polycythemia vera, seventeen with secondary polycythemia, and twelve with essential thrombocytosis were enrolled in the study, along with a control group of 22 healthy patients. DNA isolation from peripheral blood, DNA amplification with polymerase chain reaction (PCR) and agarose gel electrophoresis was performed. Using PCR-RFLP, MMP-2, -3, -9, -10, and MMP-13 gene polymorphisms were analyzed. There was a statistically significant difference between the groups in terms of Glm279Arg polymorphism rates. (respectively: v2 = 12.605, p = 0.03). There was a statistically significant difference between the groups in terms of Glm279Arg polymorphisms rates of MMP 9. (v2 = 22.975, p = 0.001). The highest MMP 9 Gln279Arg polymorphism rate was observed in the ET group. It were determined that MMP 13 gene 77A> G polymorphism had significantly higher prevalence rates in the ET group. (v2 = 12.605, p = 0.03). There was a statistically significant difference between the groups in terms of MMP-2 -735 C> T, MMP -3-1612 5A/6A, and MMP-10 180G> A polymorphism rates. (respectively: v2 = 6.447, p = 0.375, v2 = 10.870, p = 0.009). In conclusion, Glm279Arg MMP-9 and 77A> G MMP-13 gene polymorphisms had a statistically significant difference between the control and the disease groups. There was a statistically significant difference between the groups in terms of MMP-2 -735 C> T, MMP -3-1612 5A/6A, and MMP-10 180G> A polymorphism rates. (respectively: v2 = 6.447, p = 0.375, v2 = 10.870, p = 0.009). In conclusion, Glm279Arg MMP-9 and 77A> G MMP-13 gene polymorphisms had a statistically significant difference between the control and the disease groups.
Race and sex effects on the relationship between sarcopenia and BMD. H. He², Y.J. Liu³, J. Li¹, H. Shen¹, Q. Tian¹, H.W. Deng¹,². 1) Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, New Orleans, LA; 2) Center of System Biomedical Sciences, University of Shangh hai for Science and Technology, Shanghai 200093, P. R. China.

The relationship between reduced muscle mass (sarcopenia) and bone mineral density remained unclear. The main purpose of this study was to determine this relationship and examine effects of fat mass (FM), lean mass (LM), and muscle strength on regional and whole body BMD in different race cohorts and to determine if these relationships are altered by sex and/or race. The study population was collected from different clinical centers and consisted of 17,891 individuals from three ethnic populations, including 8604 Caucasians, 5013 Chinese and 4274 African-Americans. Partial correlation analyses, controlling for race, study site, gender, regular exercise, smoke and alcohol use were conducted to investigate the relationships between BMD and body composition variables. Multiple regression analysis was used to examine the independent effects of FM, LM and grip strength on regional and whole body BMD. Regression models were stratified by race and sex and adjusted for age, height, city, menopause status, current smoking, alcohol use and regular exercise. Men had a greater lean mass, lower fat mass and lower fat percentage than women. Whole body and regional BMDs were significantly greater in men than in women. African-Americans had the highest mean bone density, followed by Caucasians and then Chinese. As expected, grip strength was higher in men than in women and highest in Caucasians. Age was negatively associated with height, grip strength, lean mass, whole body and regional BMDs, and positively associated with fat mass and fat percentage. LM and FM were positively associated with BMD at all sites (r = 0.14-0.37; p < 0.001). The partial correlation was stronger for LM and BMDs than for FM and BMDs. LM was a significant independent (Std β = 0.11-0.34, p < 0.001) contributor to BMDs across race and sex and had greater effects on BMDs than FM. FM was also a significant (p < 0.001) determinant of BMD, except in African-American and Caucasian men. Compared to LM, Grip strength contributed less effect to whole body BMD and wrist ultradistal radius BMD. Non-significant race×FM and race×LM interaction terms were found. Sarcopenia was associated with a 3.473-fold higher risk of osteoporosis (OR=3.02; 95%CI=2.622, 3.473). LM is the strongest predictor of BMD for all race and sex group. Sarcopenia is associated with low BMD and osteoporosis. Further studies may assess whether maintaining lean mass contributes to prevent osteoporosis.

Unexpected pleiotropy: Do asthma and dental caries share a genetic basis? J.R. Shaffer¹, R.J. Weyant², R. Crott², D.W. McNeed², M.L. Marazita¹,². 1) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Dental Public Health and Information Management, University of Pittsburgh, Pittsburgh, PA; 3) Periodontics, West Virginia University, Morgantown, WV; 4) Dental Practice and Rural Health, West Virginia University, Morgantown, WV; 5) Oral Biology, University of Pittsburgh, Pittsburgh, PA; 6) Clinical and Translational Science, University of Pittsburgh, Pittsburgh, PA.

The surprising association between asthma and dental caries (i.e., tooth decay) has been repeatedly observed in epidemiological studies. However, at present, the cause of this relationship is largely speculative. As part of an initiative by the Center for Oral Health Research in Appalachia (COHRA), we explored the relationship between dental caries scores (assessed by intra-oral examination) and parent-reported asthma in a cohort of 940 rural children <10 years of age. These children comprise part of a larger sample of 2- and 3-generation families, all of which were genotyped on an Illumina whole-genome SNP chip. Approximately 14% of children reported asthma, which was significantly associated with both whether or not a child experienced dental caries (p=0.04; χ² test) and the number of carious (i.e., decayed or restored) tooth surfaces (p=0.001; Wilcoxon test). Using a variance components approach, we determined that both asthma and dental caries showed high heritability (90% for asthma, 40-60% for dental caries; both consistent with other studies). Moreover, estimates of genetic correlation (ρg) indicated that asthma and dental caries may partly share a common genetic basis (ρg=0.36 for asthma and number of carious surfaces, p = 0.006; ρg=0.35 for asthma and yes/no caries, p=0.03). In other words, the suites of genes affecting susceptibility to these two complex diseases may partly overlap. Next, we explored whether the ORMDL3-GSDMB locus on chromosome 17, which has been repeatedly implicated in childhood asthma in large-scale association studies, was associated with dental caries. The SNP rs8082130 in this region was associated with number of carious tooth surfaces (p=0.0004). Other loci implicated in asthma (e.g., HLA-DQA1/A2, RORA, SMAD3, and IL2RB) showed nominal associations with dental caries. These results support the hypothesis that asthma and dental caries are linked, and suggest that a shared genetic basis may partly explain this relationship. Moreover, the exact nature of this relationship remains speculative: do asthma and dental caries share a root etiology, or are asthma-liability genes influencing dental caries through a mechanistic artifact, such as exposure to antiasthma medications? While additional work is needed to tease out the causality of this association, this study illustrates that pleiotropic actions of the genetic determinants of complex diseases may yet to be fully appreciated. R01-DE014899; U01-DE018903.
A pilot T1D risk prediction study using custom panel and advanced multivariate predictive models. C. Kim1, Z. Wei, Z. Li, J. Glessner, T. Thomas1, H.K. Akerblom2, M. Knip3, H. Hakonarson4,5,6, I. Illenberger4,6,7,1. The Center for Applied Genomics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 2) The Children’s Hospital University of Helsinki, Finland; 3) The Department of Pediatrics, The Perelman School of Medicine, Philadelphia, PA; 4) Turku Centre of Genomics, University of Turku, Turku, Finland; 5) equal contribution.

Recent progress in genetic research has expanded our knowledge of the genes affecting risk of Type 1 Diabetes (T1D). Up to 50 gene loci outside the HLA region have been identified at the time of writing, and the €uro-1 study has shown that the effect of each of these genes is relatively small and their combined effect has been estimated to be only around 10-15 percent. Given the relatively modest individual contributions of T1D loci, towards genetic risk, the simple univariate methods of T1D risk prediction models have been developed. To evaluate and optimize model performance, we hypothesize that T1D can essentially be ruled out in children with a strong family history. As a proof-of-concept, here we conduct a risk prediction study for children enrolled in the full TRIGR study and early pilots. We designed a custom chip that covers 384 most relevant T1D SNPs. We genotyped 198 Finnish samples known to have pure European ancestry. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. The two types of melanin synthesized in well-defined chemical reactions are the protective dark-colored eumelanin and the sulfur-containing phaeomelanin. The events leading to the synthesis of melanin are controlled by signaling cascades that involve a host of genes encoding ligands, receptors, transcription factors, channel transporters and many other crucial molecules. The aim of this study was to evaluate polymorphisms of pigmentation genes for forensic purposes. F.T. Goncalves, F.A. Lima, R.S. Gonzales, C. Fridman, Legal Medicine, Ethics and Occupational Health, Medical School, University of Sao Paulo, Sao Paulo, Brazil.

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation traits has been the subject of intensive research in the last decade, and a number of key scientific questions are still open. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. The two types of melanin synthesized in well-defined chemical reactions are the protective dark-colored eumelanin and the sulfur-containing phaeomelanin. The events leading to the synthesis of melanin are controlled by signaling cascades that involve a host of genes encoding ligands, receptors, transcription factors, channel transporters and many other crucial molecules. The aim of this study was to evaluate polymorphisms of pigmentation genes for forensic purposes. F.T. Goncalves, F.A. Lima, R.S. Gonzales, C. Fridman, Legal Medicine, Ethics and Occupational Health, Medical School, University of Sao Paulo, Sao Paulo, Brazil.

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation traits has been the subject of intensive research in the last decade, and a number of key scientific questions are still open. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. The two types of melanin synthesized in well-defined chemical reactions are the protective dark-colored eumelanin and the sulfur-containing phaeomelanin. The events leading to the synthesis of melanin are controlled by signaling cascades that involve a host of genes encoding ligands, receptors, transcription factors, channel transporters and many other crucial molecules. The aim of this study was to evaluate polymorphisms of pigmentation genes for forensic purposes. F.T. Goncalves, F.A. Lima, R.S. Gonzales, C. Fridman, Legal Medicine, Ethics and Occupational Health, Medical School, University of Sao Paulo, Sao Paulo, Brazil.

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation traits has been the subject of intensive research in the last decade, and a number of key scientific questions are still open. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. The two types of melanin synthesized in well-defined chemical reactions are the protective dark-colored eumelanin and the sulfur-containing phaeomelanin. The events leading to the synthesis of melanin are controlled by signaling cascades that involve a host of genes encoding ligands, receptors, transcription factors, channel transporters and many other crucial molecules. The aim of this study was to evaluate polymorphisms of pigmentation genes for forensic purposes. F.T. Goncalves, F.A. Lima, R.S. Gonzales, C. Fridman, Legal Medicine, Ethics and Occupational Health, Medical School, University of Sao Paulo, Sao Paulo, Brazil.

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation traits has been the subject of intensive research in the last decade, and a number of key scientific questions are still open. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. The two types of melanin synthesized in well-defined chemical reactions are the protective dark-colored eumelanin and the sulfur-containing phaeomelanin. The events leading to the synthesis of melanin are controlled by signaling cascades that involve a host of genes encoding ligands, receptors, transcription factors, channel transporters and many other crucial molecules. The aim of this study was to evaluate polymorphisms of pigmentation genes for forensic purposes. F.T. Goncalves, F.A. Lima, R.S. Gonzales, C. Fridman, Legal Medicine, Ethics and Occupational Health, Medical School, University of Sao Paulo, Sao Paulo, Brazil.

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation traits has been the subject of intensive research in the last decade, and a number of key scientific questions are still open. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. The two types of melanin synthesized in well-defined chemical reactions are the protective dark-colored eumelanin and the sulfur-containing phaeomelanin. The events leading to the synthesis of melanin are controlled by signaling cascades that involve a host of genes encoding ligands, receptors, transcription factors, channel transporters and many other crucial molecules. The aim of this study was to evaluate polymorphisms of pigmentation genes for forensic purposes. F.T. Goncalves, F.A. Lima, R.S. Gonzales, C. Fridman, Legal Medicine, Ethics and Occupational Health, Medical School, University of Sao Paulo, Sao Paulo, Brazil.

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation traits has been the subject of intensive research in the last decade, and a number of key scientific questions are still open. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children’s Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used as a validation cohort. The two types of melanin synthesized in well-defined chemical reactions are the protective dark-colored eumelanin and the sulfur-containing phaeomelanin. The events leading to the synthesis of melanin are controlled by signaling cascades that involve a host of genes encoding ligands, receptors, transcription factors, channel transporters and many other crucial molecules. The aim of this study was to evaluate polymorphisms of pigmentation genes for forensic purposes. F.T. Goncalves, F.A. Lima, R.S. Gonzales, C. Fridman, Legal Medicine, Ethics and Occupational Health, Medical School, University of Sao Paulo, Sao Paulo, Brazil.
1171T  
**Visceral adiposity linked to chromosome 9p24.2 in adults from the Fels Longitudinal Study.** B. Towne1, J. Blangero2, A.C. Chon1, J.E. Curran1, C. Bellis2, T.D. Dyer3, E.W. Demerath1, M. Lee1, S.A. Czerninski1. 1) Wright State University School of Medicine, Dayton, OH; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) University of Minnesota, Minneapolis, MN.

Eliciting the genetic underpinnings of overweight and obesity continues to be a challenge. Part of the problem has been an over-reliance on crude measures such as weight or BMI. Using more direct and delineated phenotypic measures, however, especially of metabolically active fat depots, can provide new insights into the genetics of body composition and associated disorders. We present here results from genome-wide linkage analysis of visceral adipose tissue (VAT) volume in the entire trunk in 739 healthy adults (329 males; 410 females) aged 18 to 96 years (mean 46.1 years) in the Fels Longitudinal Study. VAT data were obtained using a multi-lice MRI protocol. Mean VAT was 3611 cc in males and 1735 cc in females. Subjects were all SNP genotyped using the Illumina Human 610-Quad BeadChip, and a subset of 17,583 SNPs identified as being in linkage equilibrium in our study population were used in whole-genome multipoint linkage analysis conducted using SOLAR (Almasy and Blangero, 1998). Significant linkage (LOD = 3.53) of VAT to markers on chromosome 9p24.2 was found, with the linkage peak being between SNPs rs7807248 and rs10758400. The 1-LOD support interval around this linkage peak contains several plausible positional candidate genes potentially relevant to visceral adiposity and cardiometabolic dysregulation, including VLDLR (very low density lipoprotein receptor), GLIS3 (GLIS family zinc finger 3), and RCL1 (RNA terminal phosphatase cyclase-like 1). Of particular interest is GLIS3 which is involved in the regulation of pancreatic beta cell development and insulin gene transcription. Recent GWAS studies have found SNPs in GLIS3 to be associated with increased risk of T1DM and T2DM. Those findings, and the results presented here, suggest that there may be a genetically mediated relationship between visceral adiposity and diabetes mellitus risk that warrants further investigation.

Supported by NIH grants R01HD012252, R01DK064391, R01DK0646870, R01AR052147, and R37MH59490.

1172F  
**Leveraging Genetic Information to Assess the Effect of Diabetes on Memory Scores in the Health and Retirement Study.** S. Walter1, J. Daniel1, L. Kubzansky2, S.C. Chang3, D.H. Rehkopf4, M.M. Glymour5, 1) Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, MA, USA; 2) Department of Medicine, Stanford University, Stanford, CA, USA; 3) Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, USA.

Background: Extensive observational analyses suggest that type II diabetes increases risk of dementia and memory impairment, but this association may be spurious and reflect unmeasured common prior causes of diabetes and dementia. The research question is not easily amenable to experimental approaches, but natural experiments based on genetic instrumental variables (IV) can lend insight into whether lifelong elevations in diabetes risk are associated with differences in late life memory functioning. Methods: Data are from 6555 non-diabetic and 1947 diabetic white participants in the nationally representative Health and Retirement Study with complete demographic, lifestyle, and health data from participants using questionnaires, medical exams, and laboratory measures; DNA is collected for a subset of participants. We selected and genotyped 63 mitochondrial SNPs to tag ancestral mitochondrial haplogroups for the major NHANES populations: non-Hispanic whites (NHW; n=8856), non-Hispanic blacks (NHB; n=4325), and Mexican Americans (MA; n=4766). A simplified mitochondrial phylogenetic tree rooted at the most recent common ancestor was constructed using these SNPs and a custom algorithm was developed to automate haplogroup classification. Preliminary tests of association were performed in NHW using linear or logistic regression to identify associations between European haplogroups and a range of phenotypes (n=132). Preliminary results revealed an association between mean cell hemoglobin and haplogroups U (p=9.0 x 10-6, b=0.12) and H (p=3.6 x 10-4, b=0.10). The correlated phenotype was mean cell volume was also associated with both haplogroup U (p=3.2 x 10-3, b=0.42) and haplogroup H (p=1.5 x 10-4, b=-0.41). The association of mitochondrial haplogroups with hemoglobin and mean cell volume is consistent with evidence that mitochondria are essential for heme synthesis. Notably, the haplogroup H-defining SNP mt10208 and the haplogroup U-defining SNP mt12308 are located in COX1 and MT-CL2, respectively. These preliminary results suggest that PhelWS is a valid approach for identifying novel mitochondrial genotype-phenotype correlations relevant to human health.

1173W  
**Employing a phenome-wide association study approach to investigate the pleiotropic nature of mitochondrial DNA variation.** S.L. Mitchell1, S.A. Pendergrass2, R. Goodloe2, K. Brown-Gentry1, R. McClanahan1, J. Boston1, M. Allen1, P. Mayo1, N. Schneitz-Boutaud4, D.G. Murdock2, D.C. Crawford1. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA.

Hematologic abnormalities, including sideroblastic anemia, have been observed comorbid with mitochondrial disorders. Rare mitochondrial (mt) DNA point mutations in COX1 and MT-CL2 have been reported in patients with anemia. Data also suggest a role for common mitochondrial variation in human disease, such as type 2 diabetes & Alzheimer’s, highlighting the potential pleiotropic nature of the mitochondrial genome. Thus far, phenome-wide association studies (PhelWS) to identify pleiotropy have focused predominantly on the nuclear genome. To characterize the pleiotropic effects of mtDNA variation, we, as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed the National Health and Nutrition Examination Surveys (NHANES), epidemiologic surveys conducted by the Centers for Disease Control and Prevention, which collect demographic, lifestyle, and health data from participants using questionnaires, medical exams, and laboratory measures; DNA is collected for a subset of participants. We selected and genotyped 63 mitochondrial SNPs to tag ancestral mitochondrial haplogroups for the major NHANES populations: non-Hispanic whites (NHW; n=8856), non-Hispanic blacks (NHB; n=4325), and Mexicano Americans (MA; n=4766). A simplified mitochondrial phylogenetic tree rooted at the most recent common ancestor was constructed using these SNPs and a custom algorithm was developed to automate haplogroup classification. Preliminary tests of association were performed in NHW using linear or logistic regression to identify associations between European haplogroups and a range of phenotypes (n=132). Preliminary results revealed an association between mean cell hemoglobin and haplogroups U (p=9.0 x 10-6, b=0.12) and H (p=3.6 x 10-4, b=0.10). The correlated phenotype was mean cell volume was also associated with both haplogroup U (p=3.2 x 10-3, b=0.42) and haplogroup H (p=1.5 x 10-4, b=-0.41). The association of mitochondrial haplogroups with hemoglobin and mean cell volume is consistent with evidence that mitochondria are essential for heme synthesis. Notably, the haplogroup H-defining SNP mt10208 and the haplogroup U-defining SNP mt12308 are located in COX1 and MT-CL2, respectively. These preliminary results suggest that PhelWS is a valid approach for identifying novel mitochondrial genotype-phenotype correlations relevant to human health.
1174T  
RNA-sequencing identifies novel differentially expressed coding and non-coding transcripts in Sjögren’s syndrome. C. J. Lessard1,2, F. Adrianto1, M. G. Dozmorov1, G. B. Wiley1, J. A. Ice1, H. Li1,2, J. A. Kelly1, A. Rasmussen1, S. B. Glenn1, K. S. Hefner1, D. U. Stone1, G. D. Houston1, D. M. Lewis1, J. A. Lessard1, J. M. Anaya5, B. M. Segal5, N. L. Rhodus1, L. Radfar2, J. B. Harley1, J. A. James1,2, C. G. Montgomery1, R. H. Scofield1,2, P. M. Gaffney1, J. D. Wren1, K. L. Swi2,1, J. B. Harley1, J. A. James1,2, C. G. Montgomery1, R. H. Scofield1,2, P. M. Gaffney1, J. D. Wren1, K. L. Swi2,1, 1) Arthritis & Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Hefner Eye Care and Optical Center, Oklahoma City, OK; 4) Valley Bone and Joint Clinic, Grand Forks, ND; 5) Universidad del Rosario, Bogota, Colombia; 6) Hennequin County Medical Center, Minneapolis, MN; 7) University of Minnesota, Minneapolis, MN; 8) Cincinnati Children’s Hospital Medical Center and the Department of Veterans Affairs Medical Center, Cincinnati, OH.

Sjögren’s syndrome (SS) is a common, clinically heterogeneous autoimmune disease characterized by exocrine gland dysfunction that involves both innate and adaptive immune responses. SS etiology is complex, with environmental, genetic, and genomic factors contributing. Many of the genetic associations reported in complex diseases, >80% map to non-protein coding DNA sequences; however, many reside in regions shown to be transcriptionally active. We used RNA-seq to identify differentially expressed (DE) protein-coding (~3% of the genome) and non-coding transcripts in 57 SS cases and 37 healthy controls. RNA samples were isolated from whole blood and prepared for sequencing using the NuGEN Encore kit and sequenced on the Illumina HiSeq 2000. Raw FASTQ files were aligned to the human genome (hg19) using TOPHAT. DE transcripts were determined using DESeq with a false discovery rate (FDR) q-value of 0.05 and a fold change of >2. After alignment, the reads were summarized for 55,076 transcripts across the human genome annotated by Ensembl. A total of 2614 DE transcripts were identified. Of the protein-coding regions, SPR14 was the most statistically DE locus in the case-control analysis (q=2.03×10E-20, fold change FC=2.32). Two other DE protein-coding transcripts of interest were identified: UCRB (q=1.94×10E-19, FC=2.86) and ATP5I (q=1.88×10E-18, FC=2.34). Among the 408 DE non-protein coding transcripts, we observed DE of a long non-coding RNA (IncRNA) at 2p25.1 (q=3.69×10E-5, FC=2.53). IncRNAs are enriched for T cell-associated genes and participate in diverse functions; however, most have yet to be characterized. Bioinformatic evaluation in the 2p25.1 region showed transcription factor binding sites and transcription of IncRNA sequences using immunologically relevant cell lines. To formulate functional hypotheses for the IncRNA at 2p25.1, we evaluated co-expression patterns with protein coding sequences and identified T cell activation and development as the most likely pathways influenced. In this SS RNA-seq study, we identified multiple candidate loci and, for the first time, DE lncRNA regions in SS. Although the function of the lncRNAs has not been elucidated, they may act as scaffolds, decoys, signals, and guides for various proteins by conferring nucleotide sequence specificity not possible by motifs alone. Future studies in SS are warranted to elucidate the functional consequences of these IncRNA.

1175F  
Genetically distinct subtypes of Rheumatoid Arthritis. B. Brynedal1, L. Klæreskog2, L. Padyukov2, L. Alfredsson1, H. Källberg1. 1) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

Complex diseases are heterogeneous in both their presentation and etiology. The identification of subgroups might therefore reveal more homogeneous populations with a smaller and distinct set of pathways. Earlier studies have demonstrated major heterogeneity regarding risk alleles in the HLA-DRB1 gene and immunologically defined (presence (ACPA+ or absence (ACPA-)) of antibodies toward citrullinated peptides) subsets of Rheumatoid Arthritis (RA). We set out to characterize how the phenotypic profile differs between subsets of RA patients and whether their genetic distinction is restricted to the HLA region. We focused our analysis on the combination of established risk factors of RA using two measures; genetic distance (GD) and genetic risk score (GRS). We used data from the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) case-control study encompassing 2747 cases (1856 ACPA+ and 891 ACPA-) and 1590 controls. The GRS was compared between groups (ACPA+, ACPA- and controls) including and excluding variants in the extended HLA region. We similarly calculated the GD between groups of RA patients and controls is significant both including and excluding HLA variants. ACPA+ and ACPA- RA were expected to be genetically distinct with a significant GD, and significantly different GRS. The exclusion of HLA variants removed much of this distinction, illustrating that RA subsets share a substantial part of their etiology. The GD between both groups of RA patients and controls is significant both including and excluding HLA variants. ACPA+ and ACPA- RA were not expected to be genetically distinct. However, the patterns suggest a more ‘protective’ HLA variant for RA patients more than controls rendering the difference significant. This also illustrates that ACPA+ RA patients carry proportionally more ‘RA protective’ HLA variants than controls. We conclude that previous published genetic risk factors show that there are both differences and similarities between subsets of RA defined by the absence or presence of ACPA. The major difference related to presence of ACPA is associated with alleles in the HLA region. Omitting the alleles in chromosome six makes ACPA+ RA and ACPA- RA more similar but different from controls. This exemplifies the importance of defining clear subsets of disease.

1176W  
Assessing human craniofacial morphology as complex phenotype in genetic association studies. F. L. Martinez1,2, J. B. Harley1, J. A. James1,2, C. G. Montgomery1, R. H. Scofield1,2, P. M. Gaffney1, J. D. Wren1, K. L. Swi2,1, 1) Anthropology Program and ICIS, Pontificia Universidad Catolica de Chile, Santiago, Chile; 2) Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, UK.

The availability of current genotyping and sequencing technologies combined with high-resolution medical imaging (such as MRIs and CT Scans) has led to a renewed interest in identifying genetic variants responsible for the genetic components of human craniofacial morphology. Recent GWAS have revealed for the first time, genetic variants correlated with normal variation in human craniofacial phenotype. Some of these loci map to genes with previously known role in craniofacial development, thus implying strong support to the findings and suggesting an unambiguous link between genotype and phenotype. There is general agreement that many genes with small effect (hundreds or thousands) should influence complex phenotypes such as body height or face morphology. Therefore, it is relevant to note that linear measurements are the sole proxy for facial phenotype currently used in GWAS. If many genes with small effect influence certain pheno- types, then the reliability of linear measurements as proxies seems an unrealistic assumption. Is it feasible to assume that linear distances account for the effect of few loci between many genes implicated in craniofacial developmental pathways? Or, is it more likely that observed variability captured by linear measurements reflects mainly the effect of non-genetic patterns of morphological variation in the population such as age or sexual dimorphism? In order to explore these questions, a small sample of 100 virtually reconstructed medical CT scans was analyzed by means of geometric morphometrics and linear measurements. In addition, single nucleotide polymorphism rs6180 (I526L) in the exon 10 of the growth hormone receptor gene (GHR) was chosen as independent variable to run correlation analysis with craniofacial morphology. This polymorphism was selected due its previous link to craniofacial morphology, human height and the role of GHR in the GH/GHF-1 axis, that promotes bone growth. A significant statistical association was found between rs6180 and craniofacial morphology. Permutation of the data was performed in order to explore the influence of sample structure on the results. Although the findings suggest a putative effect of the rs6180 polymorphism in craniofacial morphology, the correlation is actually driven by the pattern of craniofacial growth in the population. Results show that linear measurements should be regarded with more caution when used as proxy for craniofacial morphology in association studies.
SNPs and smoking: What can the aggregate of genome-wide SNPs tell us about genetic liability to smoking initiation and quantity smoked? A.G. Wilks, M.C. Keller. Institute for Behavioral Genetics, University of Colorado, Boulder, CO. Department of Psychology, University of Colorado, Boulder, CO.

Twin studies have established the genetic contribution to cigarette smoking initiation and progression and have suggested that the dimensions underlying liability to each of these stages of the smoking trajectory may be correlated. Using whole-genome SNP data from a combined sample of 10,162 European Americans that had participated in the Atherosclerosis Risk in Communities Study (ARIC) and the Multi-Ethnic Study of Atherosclerosis (MESA), we examined the degree to which the aggregate of genome-wide SNPs contributed to trait variation, or ‘SNP heritability’, for smoking initiation and quantity (cigarettes/day multiplied by years smoked). We then estimated the genetic correlation, defined as the additive genetic correlation tagged by the aggregate of SNPs, between initiation and quantity. Using the entire sample, common SNPs explained 11% (p = 0.0001) of the variance in smoking initiation and, among smokers (n = 5,967), only 4% (p = 0.2) of the variance in quantity smoked. However, when the quantity measure was analyzed separately for each sex, common genome-wide SNPs could not explain variation in quantity smoked for female smokers but explained 15% of the variance in quantity smoked among male smokers. Given that the standard error (1.2) in our female analysis of smoking quantity was large, we continued analysis on the sexes combined for the bivariate analysis of initiation and quantity. In the bivariate analysis, we re-coded initiation as an ordinal variable that included both early and late onset users in addition to those that did not initiate to allow for the estimation of the genetic correlation between initiation and quantity. When the set of SNPs that explained variance in initiation and those that impacted quantity smoked was high (r = .83) and not significantly different from 1 (p = .30). Thus, common genetic variation, as indexed by genome-wide SNPs, contributed to cigarette smoking liability, and the genetic factors that influenced smoking initiation were largely shared with those that impacted quantity smoked. Grant support: NICHD, 2T32HD007289, Stallings; NIMH, 1R01MH010141, Keller.

Endometrial expression profiling in women with recurrent early pregnancy loss, G. Kosova1, C. Billstrand1, V. J. Lynch1, M.D. Stephens2,3, C. Ober1. 1Department of Human Genetics, University of Chicago, Chicago, IL; 2Department of Obstetrics and Gynecology, University of Chicago, Chicago, IL; 3Current address: Department of Obstetrics and Gynecology, University of Illinois College of Medicine at Chicago, Chicago, IL.

Recurrent early pregnancy loss (REPL), defined as at least two documented intrauterine miscarriages of less than 10 weeks of gestation, affects up to 5% of couples trying to conceive. The endometrium’s ability to acquire a receptive state during the mid-secretory phase is a crucial component of successful implantation and, therefore, has been the focus of many studies that aim to identify these markers in order to better understand the molecular mechanisms that are perturbed in REPL. We examined the gene expression profiles of mid-secretory phase endometrium in patients with different diagnoses of endometrial dysfunction. Using the illumina HumanHT-12v4 beadChip arrays, gene expression levels were compared between women with 1) abnormally elevated cyclin E levels (n = 9) vs normal cyclin E levels (n = 23), and 2) out-of-phase histological dating (n = 10) vs normal histology (n = 22). Differentially expressed (DE) genes were identified using a likelihood ratio test within a fixed-effects linear model framework. Overall, 81 genes were DE in the first comparison and 56 in the second at P < 0.01. The two genes with the largest differential expression were ST00P (3.5-fold up-regulation in out-of-phase endometria) and LEFTY2 (4.5-fold up-regulation in out-of-phase endometria). ST00P expression peaks in the endometrium during the window of implantation (Tong et al. Fertil Steril 94:1510). LEFTY2 levels normally decrease during that phase, and aberrantly high expression levels were observed in patients with implantation failure and infertility (Tabibzadeh et al. JCEM 85:2528). Gene enrichment analysis of the DE genes (at P < 0.01) for each comparison revealed over-representation of biological processes involving immune response, signal transduction, leukocyte activation and cell adhesion among down-regulated genes in women with elevated cyclin E and cell adhesion, cell motility and ITOG-β signaling among up-regulated genes in women with out-of-phase histology. Lastly, when the analyses were limited to genes that evoluted function in the endometrium of Eutherian mammals, significant enrichments of DE genes were observed in both groups (P = 0.005 in cyclin E; P = 0.006 in ITOG-β signaling). Taken together, the results of this study suggest that this is a promising approach for identifying genes and pathways whose tight regulation during the endometrial secretory phase is necessary for optimal implantation and successful pregnancy outcome.

Association study of ERCC3 genetic variations with nasal polyposis in asthmatics, J.H. Kim1, B.L. Park2, C.S. Park3, H.D. Shin4. 1Research Institute for Basic Science, Sogang University, Seoul, South Korea; 2Department of Genetic Epidemiology, SNP Genetics, Inc., Seoul, Republic of Korea; 3Division of Allergy and Respiratory Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, Republic of Korea.

Nasal polyps, as an inflammatory condition and one of the clinical symptoms of aspirin exacerbated respiratory disease (AERD), are edematous protrusions arising mainly from the mucous membranes of the nasal sinuses. Despite the well-defined triggers, the exact mechanisms involved in the development of asthma with nasal polyps still need to be clarified. To elucidate the mechanisms involved in the progression of nasal polyps, we examined the gene expression profiles of mid-secretory phase endometrium under tight regulation during the endometrial secretory phase is necessary for that this is a promising approach for identifying genes and pathways whose genetic factors that influenced smoking initiation were largely shared with those that impacted quantity smoked. The two genes with the largest differential expression were ST00P (3.5-fold up-regulation in out-of-phase endometria) and LEFTY2 (4.5-fold up-regulation in out-of-phase endometria). ST00P expression peaks in the endometrium during the window of implantation (Tong et al. Fertil Steril 94:1510). LEFTY2 levels normally decrease during that phase, and aberrantly high expression levels were observed in patients with implantation failure and infertility (Tabibzadeh et al. JCEM 85:2528). Gene enrichment analysis of the DE genes (at P < 0.01) for each comparison revealed over-representation of biological processes involving immune response, signal transduction, leukocyte activation and cell adhesion among down-regulated genes in women with elevated cyclin E and cell adhesion, cell motility and ITOG-β signaling among up-regulated genes in women with out-of-phase histology. Lastly, when the analyses were limited to genes that evoluted function in the endometrium of Eutherian mammals, significant enrichments of DE genes were observed in both groups (P = 0.005 in cyclin E; P = 0.006 in ITOG-β signaling). Taken together, the results of this study suggest that this is a promising approach for identifying genes and pathways whose tight regulation during the endometrial secretory phase is necessary for optimal implantation and successful pregnancy outcome.

Examining ocular SNPs for regulatory enhancer potential: What does non-coding GWAS hit really mean? C.N. Simoni1, J.A. Capra2. 1Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2Department of Biomedical Informatics, Vanderbilt University, Nashville, TN.

Genome-wide association studies (GWAS) have enabled the discovery of genes and functional non-coding regions underlying many complex ocular diseases, such as age-related macular degeneration (AMD) and glaucoma. However, the mechanisms behind the vast majority of these associations remain to be elucidated. There are strong indications that many SNPs associated with ocular diseases may affect enhancer or regulatory regions. For example, two of AMD-associated SNPs in high linkage disequilibrium (LD; calculated from 1000 Genomes; r2 > 0.8) have been shown to alter a transcription factor (TF) binding site in a manner that changes gene expression. As anticipated, most ocular diseases have associated non-coding SNPs near eye developmental genes. To assess the regulatory potential of all known ocular trait-associated SNPs, we intersected them with a genome-wide set of predicted regulatory enhancers identified from analysis of histone marks, evolutionary conservation, and sequence motifs in variety of cellular contexts. Ten percent of SNPs associated with ocular traits from the NHGRI GWAS catalog fall into predicted enhancers, a slightly higher fraction than the 8.5% of all trait-associated SNPs that overlap predicted enhancers. The majority of ocular disease SNPs in predicted enhancers did not have a gene within its LD region suggesting a regulatory function. Most of the remaining predicted enhancer SNPs were in sizable introns or known TF binding sites. Comparing the genomic context of SNPs associated with different ocular traits revealed several trait-specific patterns. For example, AMD is distinct from other ocular diseases in its immunologic and cholesterol-related contributing factors. Reflecting its uniqueness, AMD was the only ocular phenotype associated with SNPs nearby complement or cholesterol genes, several of which were in predicted enhancers. GWAS hits for glaucoma are enriched in predicted enhancers when compared to other ocular diseases like AMD, pathological myopia, and corneal astigmatism. SNPs associated with glaucoma demonstrating a distinct profile suggest new functional information may in fact be gleaned from such comparative analyses. Our analysis of the regulatory potential of ocular SNPs highlights contexts and variants relevant to the associations and suggests follow-up studies. For example, profiling the expression driven by the identified enhancer variants in ocular tissues would be a useful focus for future inquiries.
A major focus of research in the post-genomic era is to decipher the heterogeneous genetic landscape underlying the pathogenesis of complex human diseases such as Autism Spectrum Disorders (ASD). A number of genes in which rare and/or common genetic variants thought to potentially play a role in ASD onset and pathogenesis have been identified. The advent of next-generation sequencing (NGS) has resulted in a significant increase in the number of ASD genes with rare genetic variants. With the accelerated growth of genetic data obtained from ASD individuals adding to the already complex genetic landscape of this disease, there is a critical need for databases specialized in the storage and assessment of this data. The autism genetic database AutDB (http://autdb.mindspec.org) was developed to serve as a publically available web-based modular database for the on-going curation and visualization of ASD candidate genes. Since its release in 2007, AutDB has become widely used by individual laboratories in the ASD research community, as well as by consortiums such as the Simons Foundation, which licenses it as STAFRI Gene. AutDB has been designed using a systems biology approach, integrating genetic information within the original Human Gene module to corresponding data in subsequent Animal Model, Protein Interaction (PIN) and Copy Number Variant (CNV) modules. The number of ASD susceptibility genes in the Human Gene Module of AutDB has increased from 304 genes in December 2011 to 546 genes in June 2013, which demonstrates both the continued discovery of ASD candidate genes and the ongoing curation of these genes into AutDB. In addition, the usage of NGS techniques has contributed to a dramatic increase in the number of rare variants identified in ASD candidate genes (from 1202 in Dec 2011 to 3061 in June 2013) compared to common variants (from 534 to 791 over the same period). Functional profiling of ASD genes with rare variants provides insight into the enriched molecular functions of these genes, including ionotropic glutamate receptor binding, serotonin receptor activity, beta-tubulin binding, and voltage-gated sodium channel activity. AutDB serves as a valuable resource for understanding the ever-evolving genetic landscape of ASD and provides researchers with information useful in bioinformatic analyses that will aid in unraveling the molecular mechanisms underlying the disease.

General financial risk factors for risk and ambiguity associated respectively with dopamine D4 receptor (DRD4) and serotonin transporter (5-HTTLPR).

A major focus of research in the post-genomic era is to decipher the heterogeneous genetic landscape underlying the pathogenesis of complex human diseases such as Autism Spectrum Disorders (ASD). A number of genes in which rare and/or common genetic variants thought to potentially play a role in ASD onset and pathogenesis have been identified. The advent of new techniques such as next generation sequencing (NGS) has resulted in a significant increase in the number of ASD genes with rare genetic variants. With the accelerated growth of genetic data obtained from ASD individuals adding to the already complex genetic landscape of this disease, there is a critical need for databases specialized in the storage and assessment of this data. The autism genetic database AutDB (http://autdb.mindspec.org) was developed to serve as a publically available web-based modular database for the on-going curation and visualization of ASD candidate genes. Since its release in 2007, AutDB has become widely used by individual laboratories in the ASD research community, as well as by consortiums such as the Simons Foundation, which licenses it as STAFRI Gene. AutDB has been designed using a systems biology approach, integrating genetic information within the original Human Gene module to corresponding data in subsequent Animal Model, Protein Interaction (PIN) and Copy Number Variant (CNV) modules. The number of ASD susceptibility genes in the Human Gene Module of AutDB has increased from 304 genes in December 2011 to 546 genes in June 2013, which demonstrates both the continued discovery of ASD candidate genes and the ongoing curation of these genes into AutDB. In addition, the usage of NGS techniques has contributed to a dramatic increase in the number of rare variants identified in ASD candidate genes (from 1202 in Dec 2011 to 3061 in June 2013) compared to common variants (from 534 to 791 over the same period). Functional profiling of ASD genes with rare variants provides insight into the enriched molecular functions of these genes, including ionotropic glutamate receptor binding, serotonin receptor activity, beta-tubulin binding, and voltage-gated sodium channel activity. AutDB serves as a valuable resource for understanding the ever-evolving genetic landscape of ASD and provides researchers with information useful in bioinformatic analyses that will aid in unraveling the molecular mechanisms underlying the disease.

1184W

HLA-DQB1 03:02 is associated with narcolepsy in the Japanese population.

1184W

Narcolepsy, a sleep disorder characterized by excessive daytime sleepiness, cataplexy, and rapid eye movement (REM) sleep abnormalities, is tightly associated with human leukocyte antigen (HLA)-DQB1 06:02. Almost all the patients carry HLA-DQB1 06:02. Since HLA-DQB1 06:02 allele is common in the general populations (10-30%), it has been suggested that it is almost necessary but not sufficient for the development of narcolepsy. To investigate the influence that additional HLA-DQB1 alleles have on susceptibility or resistance to narcolepsy, We examined HLA-DQB1 in 625 Japanese control subjects. The strongest association was with HLA-DQB1 06:01 (P = 2.2x10-10, OR = 0.37) as the second hit. Predisposing effects of HLA-DQB1 06:02 (P = 7.3x10-2, OR = 2.81) and protective effects of HLA-DQB1 06:01 (P = 2.2x10-3, OR = 0.50) were observed in narcolepsy without cataplexy, as well as with narcolepsy. No significant association was seen between idiopathic hypersomnia and HLA-DQB1 alleles. The prevalence of idiopathic hypersomnia without long sleep time (n = 82). Predisposing effects of HLA-DQB1 06:02 (P = 7.3x10-2, OR = 2.81) and protective effects of HLA-DQB1 06:01 (P = 2.2x10-3, OR = 0.50) were observed in narcolepsy without cataplexy, as well as with narcolepsy. No significant association was seen between idiopathic hypersomnia without long sleep time and HLA-DQB1 alleles. These results indicate complex HLA association with the genetic predisposition to narcolepsy and narcolepsy without cataplexy.
1185T
Association between the dopamine D4 receptor (DRD4) and political ideology in a large Singaporean Han Chinese sample. M. Monakhov1, S.H. Chew1, R.P. Ebstein2. 1) Economics Dept., National University of Singapore, Singapore; 2) Psychology Dept., National University of Singapore.

The liberalism - conservatism ideological divide has become pervasive in American politics viz., red and blue states, and understanding the determinants of this continuing conflict is of keen interest not only to political scientists and psychologists, but also to the media and the public at large. Importantly, there are marked individual differences in political ideology that have been attributed to both dispositional and situational factors. In addition to situational factors contributing to political orientation, accumulating evidence shows that social attitudes are partially heritable suggesting that genes may contribute to individual differences in liberalism - conservatism. A specific gene polymorphism, the dopamine D4 receptor (DRD4) exon III VNTR was recently identified by Fowler and his group as provisionally contributing to individual differences in political ideology, but contingent on the subject's number of friends. By leveraging our ongoing investigation of economic decision-making, we had the opportunity to undertake an independent replication of the association between DRD4 and political ideology in a large sample of ~1,800 Han Chinese university-students in Singapore. We employed the same self-report question, as did Fowler's group, to measure political ideology and demonstrate an association between the DRD4 exon III repeat region and ideology for students who grew up in a singular Singaporean political culture and are characterized by a distinctive allele distribution compared to their American counterparts. In the combined male and female sample, carriers of DRD4 exon III VNTR genotype 4R/4R reported more liberal political ideology compared to all other genotype groups (ordered logit coeff. = -0.18, p = 0.04, N = 1803). The association was evident also in an additive genetic model (regression of political attitude on number of 4R alleles in the genotype, coeff = -0.15, p=0.027). The association was observed only in females (coeff=-0.385, p=0.002, N=901) and not in male subjects. Cliff's delta (effect size measure for ordinal data) was 0.057 in combined sample and 0.11 in females subsample. Similar to the first study by Fowler's group, we also observed some evidence that the gene is associated with political ideology in women when considering the additive effect of the DRD4 on political ideology in women is mediated in part by the NEO-PI-R personality trait of Neuroticism (Sobel test: p=0.035).

1186F
An approach to complex disease modifier genes and mouse-to-human translation: transcriptome analysis of mouse strains reveals Alzheimer disease modifier gene. T. Morihara1, N. Hayashi1, H. Akatsuka2, M. Silverman3, M. Yokokoji1, N. Kimura4, M. Sato1, K. Kamino2, Y. Yamagishi2, T. Tsunoda1, T. Tanaka1, M. Takeda1. 1) Psychiatry, Osaka University, Suita, Osaka, Japan; 2) Chuo Medical Institute, Fukushima Medical Hospital, Toyohashi-shi, Aichi, Japan; 3) National Hospital Organization Hamamatsu Mental Medical Center, Nara, Japan; 4) Simon Fraser University Vancouver Cabada; 5) Laboratory of Disease Control, Tsukuba Primate Research Center, National Institute of Biomedical Innovation Tsukuba, Japan; 6) Tohoku University, Sendai, Japan; 7) Riken, Yokohama, Japan.

Background: Alzheimer's disease (AD) is a common complex disease and characterized by the accumulation of Aβeta in brain. To identify Aβeta modifier gene(s) efficiently, we postulated that combining distinct mouse strains with mosaicism would have a systematic effect and minimize the drawbacks encountered in approaches like genome-wide association studies (GWAS), human transcriptome studies and mouse quantitative trait locus (QTL) analysis. Methods: To avoid detecting secondarily affected genes by Aβeta, we used non-Tg mice in the absence of Aβeta pathology and selected candidate genes differentially expressed in the certain strain. We also prepared AD model (APP-Tg) mice with mixed genetic backgrounds. APP Tg mice were crossed onto three different strains, C57BL6 (B6), SJL and DBA/2 (DBA). The levels of Aβeta in the cortex were measured by ELISA. The levels of mRNA in the hippocampus were measured using Illumina Mouse-Ref-8 (40 arrays) and QPCR. SNPs in mice with mixed genetic background were analyzed by TaqMan assay. We also measured the total and splice variant E levels of KLC1 in human brain (14 control and 14 AD) and human fibroblasts (10 control and 10 AD control) and obtained a functional analysis of the identified gene is preformed using transfected neuroblastoma cells. Results: Aβeta accumulation levels in mice with DBA rich genetic backgrounds were drastically lower, less than one third of those in SJL or B6 rich genetic background (e.g. APP Tg mice with 75% DBA genetic background came from DBA had only 15% Aβeta40 and 32% Aβeta42 of those with 69% SJL and 31% B6 genetic background.) To identify the gene(s) which control Aβeta accumulation, we performed transcriptomics in the mouse brains. QPCR analysis confirmed a correlation between the levels of identified gene and Aβeta (Pearson’s product-moment correlation R2=0.33, p<0.0001). Genomic analysis in identified gene also supported these findings. Because the transcriptome of brain is considerably conserved between species, we continued with our transcript analysis. We examined whether the findings in mice can be translated to humans. We observed that the mRNA levels of identified gene were elevated in AD brain (+30.7%, p<0.0096).

1187W

The psychotic symptoms in Schizophrenia usually manifest in the late adolescence and early adulthood. The investigation of patients during their first episode of psychosis (FEP), particularly before they take any antipsychotic medication, could be especially helpful to disentangle the biological underpinnings of disease onset, progression and the antipsychotic effects. In this way, our aim was to carry out an expression profiling study comparing: 1) drug-naive FEP patients (n=10) and healthy subjects (n=10), and 2) drug-naive FEP patients before and after (n=10) 8 weeks of risperidone treatment. Blood samples were collected and RNA was isolated and reverse-transcribed to cDNA. To verify gene expression we used PCRarray technology, which assesses expression of 84 neurotransmitter receptor and regulatory genes plus five housekeeping genes. We considered as a detectable level of expression those genes with ΔCt < 35 - HKG (housekeeping genes) geometric mean of each group sample. We were able to identify three groups of genes: A, genes not expressed in any group (14 genes); B, genes not expressed in some groups or lesion group (10 genes); and C, genes expressed in all groups (55 genes). For group C, we performed Mann-Whitney U test (FEP×Control) or the Wilcoxon test (FEP×Before×After) and Bonferroni correction for multiple comparisons, and for group B we used Fisher exact test, in order to find genes that are expressed only under a specific condition. After Bonferroni correction, none of the associations remained for the group C of genes (p > 0.009). Concerning group B, we found that PROKR1 expressed preferentially in drug-naive FEP patients compared to the FEP patients after treatment (p=0.02), and with a tendency between FEP and controls (p=0.07). Our results suggest that treatment with risperidone may decrease expression levels of PROKR1, returning to undeectable expression in FEP patients after treatment such as the control group. However, this gene has not been investigated in psychiatric disorders, and further studies are needed to investigate the role of the prokinecins in antipsychotic treatment. These present data are relevant in terms of cost and effectiveness to studies of gene expression. PROKR1 was expressed only in FEP group, but not in others, indicating that its expression could be regulated by the psychosis condition or risperidone treatment.
1188T
The cerebral glucose transporter SLC45A1 is mutated in individuals with non-syndromic intellectual disability and epilepsy. M. Srou1,3, N. Shimokawa2, F.F. Hamdan1, S. Dobczyniecka2, G.A. Rouleau2, C. Poulin2, J.L. Michaud1, 1) Sainte-Justine Hospital Research Center, Montreal, H3T 1C5, Canada; 2) Division of Pediatric Neurology, Montreal Children’s Hospital, McGill University Health Center, Montreal, H3H 1P3, Canada; 3) Department of Molecular Neurosciences, University of Montreal. The exome dataset was analyzed by looking for homozygous rare (<0.5% minor allele frequency) coding and splicing variants found in the candidate homozygous regions. We excluded variants that did not segregate with the phenotype in the family or that were found in 95 ethnically-matched controls. Of the remaining variants, only one missense variant in SLC45A1 (c.C629T/p.A210V) was predicted damaging by Polyphen-2 and mutation taster. SLC45A1 encodes a glucose transporter that has only recently been characterized. Slc45a1 is highly expressed in the brain of mice, is induced after hypercapnia and mediates glucose uptake along the pH gradient. We show that Slc45a1-transfected COS-7 cells carrying the same c.629T (p.A210V) mutation exhibited a 40-50% decrease in intracellular glucose uptake, suggesting that the mutation affects the function of the Slc45a1 transporter. Mutations in another cerebral glucose transporter, GLUT1, are implicated in neurologic disease and are well known to result in epilepsy with variable degrees of intellectual disability, or movement disorders. Our results suggest the possibility of specific treatment in the form of a ketogenic diet for the family, and implicate for the first time SLC45A1 in human identification. The addition of similarly affected families will be necessary to establish SLC45A1 as a neurodevelopmental disease gene. Collaborations towards this effort are welcome.

1189F
Adducin function is essential for sustained changes in neuronal activity upon learning. V. Vukojevic1, F. Peter2, P. Demougin1, N. Hadziselimovic1, J.-F. de Quervain1, F.F. Hamdan1, S. Dobczyniecka2, G.A. Rouleau2, C. Poulin2, 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 2) University of Basel, Department Neurobiology, Life Sciences Training Facility, Klingelbergstrasse 50/70, 4056 Basel, Switzerland; 3) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 4) University of Basel, University Psychiatric Clinics, Wilhelm Klein-Strasse 27, 4055 Basel, Switzerland. Adducin, a member of the protein family of actin-capping activity (Vukojevic et al., 2012). In the present work, we further investigated the effects of adducin dependent changes on neuronal activity during synaptic transmission. With the help of genetically encoded calcium indicators (GECIs) we visualized the neuronal activity in vivo in defined neuronal populations. Specifically, we monitored calcium currents in neurons of adducin knock down mice. We found that adducin knockdown mice failed to show activity after 30 minutes. Therefore we were also able to investigate in vivo the role of neurons in memory storage. These findings further support the role of ADD-1 in the stabilization of synapses, changes in GLR-1 dynamics and finally modulations of neuronal activity patterns. Together, all of these results suggest that the lack of adducin in AVA interneurons leads to an abnormal synaptic remodelling and changes of neuronal activity that are functionally reflected also in other members of the motor network.

1190W
Influence of Alzheimer’s disease genes on cognitive decline: the Guangzhou Biobank Cohort Study. S.S. Cherry1, H.S. Gu1, L. Xu2, P.C. Shun1, C.-O. Jiang2, T.H. Lam3, B. Liu4, Y.L. Lin5, T. Zhu1, W.S. Zhang1, G.N. Thomas2, K.K. Cheng3, 1) Psychiatry & Centre for Genomic Sciences, Univ Hong Kong, Pokfulam, Hong Kong; 2) Department of Community Medicine, School of Public Health, University of Hong Kong, Hong Kong; 3) Guangzhou No. 12 Hospital, Guangzhou, China; 4) Public Health, Epidemiology, and Biostatistics, University of Birmingham, Birmingham, UK. Mild cognitive impairment is a reliable predictor of the future onset of clinical dementia. In Hong Kong, the prevalences of very mild dementia and mild dementia in people aged 70 years and older were estimated to be 8.5% and 8.9%, respectively, in a population-based sample. For those subjects with a clinical diagnosis of dementia, Alzheimer’s disease (AD) was the most common likely cause (73.5%), with 22.4% having vascular dementia (VaD), and 3.9% dementia with symptoms of Parkinson’s disease. In addition to the impact on the patient and their family members, the economic cost of AD has been estimated to range from around $48,000 to $585,000 per patient per year around the world. To identify risk factors involved in cognitive decline, we selected 1325 extreme cognitive decline subjects and 1083 matched controls from Guangzhou Biobank Cohort study (GBCS) for DNA genotyping at 30 known AD-associated SNPs. Full information maximum likelihood (FIML) regression was adopted to analyse quantitative change scores, while multiple logistic regression was used to investigate the genetic effect of those variants on dichotomous phenotype. No allelic association was found by individual variant analysis. At the level of genotypic association, not only we did confirm that the APOE ε4 homoyzygote can significantly predict cognitive decline (p < 0.05), but also revealed carriers of the ACE rs1800764 C allele are more likely suffer decline than non-carriers, especially in the samples without college education. However, these effects together only explain 1.3% of the phenotypic variance, and suggest that AD risk variants/genes are only minor predictors of cognitive decline in these Chinese samples.

1191T
Fine Mapping and Association Analysis of Candidate Genes for Autism Spectrum Disorder and Language Impairment in the NJLAGS Sample. A. Hare1, A. Seto1, J. Fleix1, M. Azrano1, S. Byskse2, C. Bartlett2, Z. Bzowski3, 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Department of Statistics, Rutgers University, Piscataway, NJ; 3) Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children’s Hospital and The Ohio State University, Columbus, OH. Over the past decade, the New Jersey Language and Autism Study (NJLAGS) has collected families that contain one individual with autism and another individual with specific Language Impairment (SLI) but not autism. This is the first study of its kind to investigate the share genetics between autism and SLI. Using a comprehensive neuropsychological testing battery, three categorical phenotypes: language impairment (LI), reading impairment (RI), and social impairment (SRS-DT), and two quantitative phenotypes: autism proband scores were included in the quantitative cognitive decline, we selected 1325 extreme cognitive decline subjects and 1083 matched controls from Guangzhou Biobank Cohort study (GBCS) for DNA genotyping at 30 known AD-associated SNPs. Full information maximum likelihood (FIML) regression was adopted to analyse quantitative change scores, while multiple logistic regression was used to investigate the genetic effect of those variants on dichotomous phenotype. No allelic association was found by individual variant analysis. At the level of genotypic association, not only we did confirm that the APOE ε4 homoyzygote can significantly predict cognitive decline (p < 0.05), but also revealed carriers of the ACE rs1800764 C allele are more likely suffer decline than non-carriers, especially in the samples without college education. However, these effects together only explain 1.3% of the phenotypic variance, and suggest that AD risk variants/genes are only minor predictors of cognitive decline in these Chinese samples.
TRPM8 variants that protect from migraine make individuals less susceptible to cold pain. M. Kaunisto1,2, E.J. Holmström2, V. Anttila1,3,4,5, M. Kallol6, E. Hämäläinen1,5, A. Stubhaug1, M. Wessman1,2, C.S. Nielsen1, E. Kalso1,5,10, A. Palotie1,4,6, E. Kalso1,5,10, A. Palotie1,4,6, E. Kalso1,5,10, A. Palotie1,4,6. 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Folkhälsoan Research Center, Helsinki, Finland; 3) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Wellcome Trust Sanger Institute, Welcome Trust Genome Campus, Hinxton, United Kingdom; 6) Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland; 7) Division of Emergencies and Critical Care, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 8) Division of Mental Health, Norwegian Institute of Public Health, Oslo, Norway; 9) Institute of Clinical Medicine, University of Helsinki, Helsinki, Finland; 10) Pain Clinic, Department of Anaesthesia and Intensive Care Medicine, Helsinki University Central Hospital, Finland.

Several common variants predisposing to migraine have recently been identified. The most consistently associating variants lie within or around TRPM8. TRPM8 is a ligand-gated cation channel that plays a key role in sensing cold pain. Although the function of this channel has been intensively studied, no evidence of TRPM8 variants affecting cold pain sensitivity exists. To elucidate the putative role of TRPM8 in cold pain, we genotyped two migraine-associated SNPs, rs10166942 and rs17862920, in two study cohorts. First, of these, the BrePainGen study cohort consists of 1,000 Finnish patients undergoing breast cancer surgery and tested for experimental pain before surgery. From the second cohort, the population-based Tromsø study, 2,369 Norwegian individuals tested for experimental cold pain were studied. Cold pain was measured in both cohorts by immersing the hand to circulating +3°C water for the maximum time tolerated by the study subjects. Pain intensity was assessed at regular intervals with a 0 to 10 Numerical Rating Scale (NRS). Association between TRPM8 SNPs and NRS values as well as time to withdrawal was examined with linear regression analysis. Sex, age, BMI and anxiety status were included as covariates. The primary phenotype was the NRS value reported soon after the beginning of the test (at 15 s or 13 s, depending on the cohort).

Meta-analysis combining the results of these two cohorts showed that both of the SNPs were associated with cold pain sensitivity at the specified time-point (p=0.001, beta=0.24 for rs10166942 and p=0.004, beta=0.3 for rs17862920, assuming an additive model). The carriers of the minor alleles reported less intense cold pain than the non-carriers which is in line with the fact that the minor alleles of these SNPs are known to protect from migraine. The mean NRS of the homozygous rs17862920 T-allele carriers was 1.2 scores lower, compared to the rest of the subjects.

When the later time-points were examined, it became evident that the effect was only seen in the beginning of the test. Furthermore, there was no evidence of association between the SNPs and cold pain tolerance. This suggests that the difference is in the early sensory processing of cold pain. Based on these results it seems possible that rather than being migraine specific susceptibility variants these SNPs may have a role in more general pain sensitivity mechanisms.

The Role of CNTNAP2 in a Developmental Language Disorder and Potential Mechanisms. N. Li1, S.L. Wolock1, S.A. Pettif1, J.F. Flax2, A.S. Rascher3, L.M. Brzustowicz3, C.W. Bartlett1. 1) The Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children’s Hospital and Department of Pediatrics, The Ohio State University, Columbus, OH; 2) Department of Psychology, The Ohio State University, Columbus, OH; 3) Genetics, Rhode Island Hospital, Providence, RI; 4) Department of Psychiatry, University of Toronto, Toronto, CA.

Background: The purpose of this study was to assess CNTNAP2 in a sample of extended pedigrees ascertained for specific language impairment (SLI). SLI is a development disorder in which children learn language at a rate slower than the expected rate despite normal cognition, speech patterns and hearing. CNTNAP2 is associated with different behavioral disorders including SLI and autism, as well as with normal language development. However, studies that have focused on SNPs in CNTNAP2 have not focused on the gene in the context of functional potential. Here, we performed three studies to examine the potential function of CNTNAP2 in language development. Methods: STUDY 1: We used dense SNPs in 21 nuclear and extended SLI families from the U.S. and Canada. Nine language and reading phenotypes were tested for linkage/association. STUDY 2: We conducted a meta-analysis of 798 subjects from five published human brain expression Quantitative Trait Loci (eQTL) studies with SNPs imputed to the 1000 Genomes reference haplotypes. STUDY 3: We performed follow-up analysis of SNPs in 106 human brain samples (frontal lobe) from our lab. Results: STUDY 1: Using the quantitative trait analysis of a standardized language assessment, we observed a multipoint posterior probability of linkage (PPL) of 85% (MOD=4.7). Analysis of a categorical written language impairment diagnosis yielded a PPL of 40%, and similar results were observed with the other outcomes. No significant association was found with autism or schizophrenia. STUDY 3: Follow-up analysis using qPCR on an in-house collection of brain samples also indicated three SNPs that were potentially eQTLs for CNTNAP2. Conclusions: While CNTNAP2 is associated with multiple neurocognitive phenotypes, there is no evidence that CNTNAP2 influences general intelligence. Furthermore, the peak linkage in our samples appears to be driven by one or more eQTLs within CNTNAP2 that influence expression in a genotype dependent way. Studies to isolate the causal SNP(s) that actually influence expression levels are warranted.

Association of Language Impairment with Language, Reading, and ADHD candidate genes. S. Smith1, J. Gayan1, M. Rice1. 1) Dept Pediatrics, Univ Nebraska Med Ctr, Omaha, NE; 2) Bioinfosal, Sevilla Area, Spain; 3) Dept. Speech, Language, Hearing, Univ Kansas, Lawrence, KS.

Specific language impairment is a significant developmental disorder in language acquisition occurring in about 7% of 6-year-olds in the US. Estimates of penetrance is a disorder based on the severity, with heritabilities greater than 0.5 for children who had been seen by a speech pathologist. We previously reported association of LI with SNPs in a Reading Disability candidate gene, KIAA0319, and with a language candidate gene, FOXP2. The current study expands the coverage of these and three other known genes, with concentration on SNPs in regulatory regions. We genotyped 422 SNPs in 23 candidate genes in 293 people from 62 nuclear families with a proband with SLI. Probands were between 3 and 12 years old and were ascertained through school-based speech pathologists, with diagnosis confirmed with standardized testing appropriate for the child’s age. Their siblings between ages 2 and 17 were also recruited and tested. All children were given an additional battery of 9 tests covering reading and language abilities. Quantitative association analysis was done with the PBAT program implemented by SVS (Golden Helix) testing the hypothesis of association in the presence of linkage since these were candidate genes with some previous evidence for relationship with LI or a related disorder. Several genes and phenotypes showed association with multiple SNPs. The strongest evidence for association was in the expected genes FOXP2, CNTNAP2, and KIAA0319, but additional evidence was found with DRD2, BDNF, and DBH. The strongest results were with FOXP2. The most significant SNPs were concentrated at the 5’ end of the gene, with the haplotype rs907533.rs877534.rs7371635 (C:T:G) at a nominal p=0.0004 for the Gray Oral Reading phenotype. For CNTNAP2, the haplotype rs12703865.rs9840492.rs10952672 (A:G:A) showed a nominal p=0.0013 for a measure of vocabulary growth. For KIAA0319, the strongest haplotype was not in the same region that is generally associated with reading: rs807533.rs2760179.rs807509 (G:G:C) p=0.0008 for a reading phenotype, however, another haplotype was located within the putative reading disability region: rs3067943.rs2294869.rs1431340 (G:C:T) with the Gray Oral reading phenotype, p=0.0007. We also replicated association with PCDH9 on chromosome 13 with 3 SNPs, all with the TEGI grammar phenotype: rs2324908 (p=0.0084), rs4883774 (p=0.0094), and rs9564308 (p=0.0094). An additional 21 families has been genotyped on the same platform and will provide additional data.
1195F
Molecular delineation of the 1p36.3 locus and candidate gene resequencing in 363 patients with polymicrogyria. S. Beijnsberger1, G.L. Carvill1, S. Berkoek2, I.E. Scheffer2, W.B. Dobyns2, H.C. Meford1, 1) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, WA 98195, USA; 2) Epilepsy Research Center and Department of Medicine, University of Melbourne, Austin Health, Australia; 3) Pediatric and Neurology, University of Washington, Seattle, WA 98195, USA.

Polymicrogyria (PMG) is characterized by an abnormal folding pattern of the cortex, which results in an excessive number of gyri and abnormal cortical lamination. PMG occurs in several forms and causes a wide spectrum of phenotypes, ranging from minimal neurologic manifestations to severe intellectual disability and refractory epilepsy. Mutations in a small number of genes have been shown to cause PMG, including TUBA1A, TUBB5, TUBB2B, TUBB3 and WDR62. Furthermore, copy number variants (CNVs) at several genomic loci have been consistently linked to PMG, including the 1p36.3 locus. The 1p36.3 critical interval for PMG was previously localized to the terminal 4.8Mb of 1p. Using genome-wide array-CGH we recently detected an interstitial 4Mb de novo 1p36.3 deletion in a patient with bilateral perisylvian PMG. Importantly, this deletion partially overlaps with the known 1p36.3 critical interval and allowed us to refine this PMG locus to a 2.3Mb region (chr1: 2.5-4.8Mb) encompassing 12 genes. We selected six candidate genes from this region for targeted resequencing, based on gene function and expression in the brain: ACTR72, AJAP1, ARHGEF16, TP73, TPRG31L and WRAP73. Using molecular inversion probes for targeted capture followed by multiplexed next generation sequencing we are screening 363 patients with PMG, for mutations in these genes. To date we have sequenced 118/363 probands with 90% of the target base pairs sequenced at >25X coverage, and identified non-synonymous variants in four candidate genes (TP73, WRAP73, ARHGEF16 and AJAP1) that are not present in 6500 control exomes. Segregation analysis of these variants to determine whether these variants arose de novo and are therefore likely causative for PMG is underway. We have initiated this targeted resequencing approach and segregation analysis will be performed for the remaining 245 probands to identify de novo mutations in our candidate genes. Furthermore, we are expanding our targeted resequencing approach to include candidate genes from additional CNV loci and CNVs identified in our sequencing studies. In conclusion, we are expanding our targeted resequencing approach and segregation analysis will be performed for the remaining 245 probands to identify de novo mutations in our candidate genes. Furthermore, we are expanding our targeted resequencing approach to include candidate genes from additional CNV loci and CNVs identified in our sequencing studies.

1195T
FOXG1 is implicated in abnormal brain development in two unrelated cases with 14q aberrations. G. Macintyre1, K. Schlade-Bartusiak1, M.L. Martinez-Fernández2, M.L. Martinez-Frias2, R. Arteaga3, P. O’Brien4, M.A. Ferguson-Smith2, D.W. Cox2. 1) Medical Genetics, University of Alberta, Edmonton, Alberta, Canada; 2) Research Centre for Congenital Anomalies (CIAC), CIBER de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain; 3) Neuropediatrics service, Hospital Universitario Marques de Valdecilla, Santander, Spain; 4) Centre for Veterinary Science, University of Cambridge, Cambridge, UK.

We have characterized two cases of de novo 14q rearrangements on the proximal long arm of chromosome 14, each associated with abnormal brain development. The first is a de novo inversion in a male, who presented at 6 months with microcephaly, hypotonia, epilepsy, developmental and psychomotor delays, and partial agenesis of the corpus callosum. Weight and length were normal. At 11 months, the patient exhibited partial, generalised seizures. At 20 months, developmental delay was evident, weight 10.5Kg (25th centile), length 81cm (25th centile) and occipital frontal circumference (OFC) = 43cm (<3rd centile). Karyotype and FISH analyses revealed a de novo paracentric inversion: 46,XY,inv(14)(q12q21.3). The proximal break lies between genes, while the distal break is in a gene-rich area and may cause a gene disruption. Candidate genes for disruption include KLHDC1, KLHDC2, SOCCAG1 and ARF6. However, lymphoblast gene expression analyses indicated altered expression of several other genes compared to controls, including genes adjacent to the proximal breakpoint at 14q12: FOXG1, position unchanged, is upstream of the proximal breakpoint and is up-regulated in patient lymphoblasts, while ARHGPAS5 is down-regulated. FOXG1, typically known as Brain Factor 1 (BF1), is essential to normal mammalian brain development and FOXG1 gene defects have been associated with corpus callosum defects, as well as an atypical Rett syndrome. Of additional interest, is the altered regulation of ARHGAP5, encoding RhoGAPSp190B, as p190B defects have previously been associated with altered neuronal migration, as well as major forebrain and corpus callosum defects in mice. The second case involves the posthumous CGH characterization of a 4.2Mb deletion within 14q12 (46,XY,del(14)(q11q13)). The patient exhibited microphthalmia, webbed neck, syndactyly, and corpus callosum agenesis of the corpus callosum. Several genes were found to be deleted including NOVA1, FOXG1, and PRKD1. Our data from two unrelated cases support published findings from various authors that identify the transcription factor, FOXG1, as a protein essential for the normal development and function of the brain.

1196W

The purpose of the project is to perform higher-resolution genotyping by aCGH on subjects with autism and ADHD to determine the prevalence of CNVs in genes within the metabotropic glutamate receptor (mGluR) network. Recently, Center for Applied Genomics/CAG has found genes in this network were enriched for CNVs in patients with ADHD and autism (data not published). The analysis was based on genotyping on Illumina SNP arrays (550–610K). The CNV analyses revealed that ~10% of individuals with ADHD and ~2–3% of individuals with autism have CNVs within genes in the mGluR network. However, the resolution of the analyses was limited by the spacing of the probes in these SNP arrays (~4–6kb between probes) and potentially important CNVs that were missed due to lack of sensitivity of the arrays. Therefore, we designed a custom array for comparative genomic hybridization (CGH) with dense coverage (~300–400bp between probes) within genes in the mGluR network as well as other genes implicated in ADHD and other mental disorders. The ADHD and autism cohort within CHOP that were used in the previous analyses will be systematically screened again with the custom CGH array for mGluR network mutations. Importantly, those subjects with these mGluR network mutations could represent a distinct subset where the etiology of their disorder arises from mGluR dysfunction. This potential discovery could lead to responsive therapeutic with mGluR modulators.
1198F
Uncovering Obsessive-Compulsive Disorder risk genes using a high-resolution genome-wide CNV approach. M.J. Gazzellone1,2, S.M. Shaheen2, B. Li3, R.J. Schachar1, A.C. Lionel3,4, B. Thiruvahindrapuram1, M. Udden5, K.D. Fitzgerald6, C.R. Marshall7, D.R. Rosenberg5,6, N. Soreni7, G.L. Hanna8, P.D. Arnold1,8, S.W. Scherer1,4, *1) The Centre for Applied Genomics and Program in Genetics and Genomic Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 3) Department of Psychiatry and Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Psychiatry, University of Michigan Medical School, Ann Arbor, MI, United States; 5) Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI, United States; 6) The Children’s Hospital of Michigan, Detroit, MI, United States; 7) Department of Psychiatry and Behavioural Neurosciences, Faculty of Health Sciences, McMaster University, St. Joseph’s Healthcare, Hamilton, Ontario, Canada; 8) Department of Psychiatry and Institute of Medical Science, Toronto, Ontario, Canada.

Background: Obsessive-Compulsive Disorder (OCD) is a neuropsychiatric condition characterized by persistent unwanted thoughts (obsessions), ritualized actions and repetitive behaviors (compulsions), and excessive anxiety. It is common (2% lifetime prevalence) and often presents during childhood. Family and twin studies suggest that genetic factors underlie the pathology of OCD, particularly when the symptoms begin in childhood. To date, genetic investigations have focused primarily on candidate gene, genetic linkage, and genome-wide association approaches, but there have been no published copy number variation (CNV) studies. Findings reported to date explain only a fraction of the genetic architecture of OCD. Our approach seeks to build upon the foundation established in studies of other neuropsychiatric conditions which posit that rare inherited and de novo CNVs may elevate disorder risk. We aimed to establish whether this type of variation contributes to the pathogenesis of OCD.

Methods: We undertook a genome-wide CNV scan using two high-resolution microarrays: the Affymetrix CytoScan HD array and the Illumina Omni 2.5M Quad array (both featuring around 2.5 million probes). In stage one of the project, DNA was obtained from 100 affected children and their parents and run on the CytoScan HD array. In stage two, we added another 73 children and their parents to run on either the CytoScan HD array (50 probands) or the Omni 2.5M Quad array (112 probands). We identified high-confidence CNVs by requiring their identification by two or more CNV detection algorithms. Rare variants were identified by comparing the stringent CNV calls from our case cohorts to stringent calls obtained from population-based controls genotyped on the same array (873 CytoScan HD controls and 2,988 Omni 2.5M Quad controls).

Results: Rare CNVs overlapping genes previously implicated in neuropsychiatric disorders, and absent in the controls, were uncovered in the OCD probands. These include exonic deletions or duplications of genes involved in neuronal migration (i.e. ASTN2), synaptic function and signal transmission (i.e. NLGN1 and NLGN4X), and postsynaptic scaffolding in glutamatergic synapses (i.e. DLGAP1 and DLGAP2). In addition, we identified new candidate OCD genes involved in synaptic function or plasticity (i.e. BTBD9). Our findings suggest that rare copy number changes contribute to OCD risk and that genes expressed at the synapse may play a role in the onset of OCD when perturbed.

1199F
Family structure is predictive of genetic architecture in autism spectrum disorders. K. Schmitz-Abel1,2,4,5, 6, M. Chahroudi1,2,3,4,5, 6, T. Wirth1,2,3,4,5, 6, S. Hille1,2,4,5, 6, G. Sanchez-Schmitz1,2,4,5, 6, J. Parfitt1,2,3,4,5, 6, B. Barry1,2,3,4,5, 6, B. Mehta1, S. Servattelab1, A. Ngoc Lam1, C. Walsh1,2,3,4,5, 6, K. Markianos1,2,3, K. McKenzie1,2,3,4,5, 6, 1) Division of Genetics, Department of Medicine, Children’s Hospital Boston, Boston, Massachusetts, USA; 2) Manton Center for Orphan Disease Research, Children’s Hospital Boston, Boston, Massachusetts, USA; 3) Howard Hughes Medical Institute, Children’s Hospital Boston, Boston, Massachusetts, USA; 4) The Autism Consortium, Boston, Massachusetts, USA; 5) Departments of Pediatrics and Neurology, Harvard Medical School, Boston, Massachusetts, USA; 6) Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA; 7) Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; 8) The Division of Infectious Diseases, Boston Children’s Hospital; 9) Harvard Medical School, Harvard University Boston, MA 02115, USA.

Analyses of large autism datasets have provided statistical and functional evidence for the role of rare point mutations and transmitted and de novo copy number variants (CNVs), and offer crucial insights into the diverse genetic mechanisms that can lead to Autism Spectrum Disorders (ASDs). Here we present CNV analysis for a cohort of 183 consanguineous families with one or more children affected with ASD. We provide new insights into the genetic architecture of ASDs as our cohort is uniquely enriched for recessive loss of function variants. We follow up findings and draw comparisons with additional large ASD and control datasets: the Simons Simplex and the Autism Genetic Resource Exchange (AGRE) collections (2,670 affected individuals; 9681 total individuals). Comparing across these cohorts, we find that de novo CNVs are more common in non-consanguineous families, with a single affected child (p=0.04), but a lesser role in multiplex families, and they are no more common in ASD cases than controls in multiplex consanguineous families. In contrast, we present the strongest statistical evidence to date that homzygous deletions in consanguineous families are a major contributor to ASD disease burden in consanguineous families, contributing to as much as 5–10% of cases.

1200T
Susceptibility Loci for Pigmentation and Melanoma In Relation To Parkinson’s Disease. J. Dong1, J. Gao2, M. Nalls3, X. Gao4, X. Huang4, J. Han5, A. Singleton6,9, IPDGC. IPDGC, 8, H. Chen6, 7) NIEHS, durham, NC; 2) National Institute on Aging, Bethesda, Maryland; 3) Harvard School of Public Health, Boston, Massachusetts; 4) Pennsylvania State University-Milton S. Hershey Medical Center, Hershey, Pennsylvania; 5) Channing Laboratory, Harvard Medical School, Boston, Massachusetts; 6) Internation Parkinson’s Disease Genomics Consortium (IPDGC).

Objective: To examine potential associations between genetic susceptibility loci for pigmentation or melanoma and the risk for Parkinson’s disease (PD). Methods: We examined 13 SNPs from previous GWAS studies on pigmentation or melanoma in relation to PD risk among 808 PD cases and 1,623 controls in the Parkinson’s Genes and Environment (PAGE) Study and then a broader selection of 360 SNPs among 5,333 cases and 12,019 controls from the International Parkinson’s disease Genomic Consortium (IPDGC). In the PAGE study, we also examined the colors of hair, eye, or skin and melanoma in relation to PD risk. All participants were self-reported Caucasians. Results: As expected, many of these SNPs were associated with pigmentation or melanoma. However, neither these SNPs nor pigmentation phenotypes were associated with PD risk after adjusting for multiple comparisons. In PAGE, 18 PD cases (2.2%) and 26 (1.6%) controls also had melanoma, resulting in an odds ratio of 1.3 (95% confidence interval: 0.7–2.4). The IPDGC analysis confirmed that none of the pigmentation or melanoma SNPs were associated with PD risk. Conclusions: Our study did not show any association between pigmentation or melanoma related SNPs and the risk for PD.
1201F
Genetic Risk Factors in Utah Pedigrees at High Risk for Suicide. H. Coon1, R. Pimentel2, K.R. Smith2, C. Huff3, H. Hu4, L. Jeronimskij, J. Hansen1, M. Klein1, W.B. Callo4, J. Byrd5, A. Bakian1, S. Crowell1,2, W.H. McManus1, E. McGlade1, D. Yurgelun-Todd1, P. Renshaw1, T. Grey2, D. Gray1. 1) Dept. Psychiatry, Univ Utah, Salt Lake City, UT; 2) Pedigree & Population Resource, Huntsman Cancer Inst., Univ Utah, Salt Lake City, UT; 3) Dept Family Consumer Studies, College Soc Behav Sci, Univ Utah; Salt Lake City, UT; 4) Dept Epidemiol MD Anderson Cancer Center, Univ Texas, Houston, TX; 5) Core Research Facilities, Health Sci Center, Univ Utah, Salt Lake City, UT; 6) UT State Office Medical Examiner, UT Dept Health, Salt Lake City, UT; 7) Dept Psychology, Univ UT, Sale Lake City, UT.

We used unique population-based data resources to identify 22 high-risk extended pedigrees that show clustering of suicide twice to three times that expected from sex- and cohort adjusted incidence rates. We then conducted a genome-wide association study using genetic data from the above pedigrees as well as seven previously published studies. These results showed the Met allele of the BDNF Val66Met SNP tended to associate with suicide in the Japanese population but not with the completed suicide subjects.

1202W
Association study of BDNF with completed suicide in the Japanese population. A. Hishimoto1,2, W. Ratta-apa1, K. Shiroiwa1, I. Sora1. 1) Dept. of Psychiatry, Kobe Univ Grad Sch Med, Kobe, Japan; 2) Dept of Psychiatry and Behavioral Science Albert Einstein College of Medicine, Bronx, NY, USA.

Brain-derived neurotrophic factor (BDNF) is involved in neuronal survival, brain plasticity, and neuronal development. Impairment of BDNF has been implicated in the pathophysiology of psychiatric disorders and suicide. We prioritized variants that were: 1) shared across pedigree members, 2) rare in publicly-available sequence data from 1,358 controls, and 3) screened against 258 other Utah suicides not in the pedigrees to eliminate potential false positives. Sequence variants were prioritized statistically and then implicated genes were screened for previous disease associations and functional relevance. Findings included membrane protein genes and several genes with neuronal involvement and/or known associations with psychiatric conditions. Gene implicated in particular pedigrees may be associated with significant co-morbid psychiatric or medical conditions and/or demographic attributes unique to that pedigree. While the study is limited to variants identified on the HumanExome BeadChip, these findings warrant further exploration, and demonstrate the utility of this high-risk pedigree resource.

1203T
Rare and common variants near CHRNB3-CHRNA6 are associated with cocaine dependence. B. Sadler1,2, G. Haller3, N. Saccone1, L. Bien1, A. Goate1. 1) Study of Addiction, Genetics and Environment (SAGE) Collaborators. In the U.S.A., cocaine is the second most frequently abused illicit drug. Although there have not yet been large-scale GWAS for cocaine dependence, the CHRNA3-A5-B4 nicotinic receptor gene cluster on chromosome 15, identified by GWAS to be associated with nicotine dependence, also harbors variants associated with decreased risk for cocaine dependence. Variants within the CHRNA6-B3 gene cluster on chromosome 8 were found to affect nicotine consumption in several GWAS and thus represent intriguing candidate genes for the study of drug dependence. Using genotype data from a GWAS of the Study of Addiction: Genetics and Environment (SAGE) dataset including 2714 European Americans (605 subjects with non-zero scores on the DSM-IV cocaine symptom count and 2109 subjects with scores of zero on the DSM-IV cocaine symptom count), we tested for association of CHRNA6-B3 SNPs with DSM-IV cocaine dependence symptom count. The top SNP for symptom count was rs9298626 (β = 0.528, p=7.0×10−4), although multiple SNPs passed the multiple test correction. To determine whether there was evidence for multiple independently associated variants at the locus contributing to risk for this phenotype, the most significant SNP was added as a covariate in a linear regression model. We find evidence for more than one independent signal within this locus in European Americans for cocaine dependence symptom count. Further, when using only exposed but not non-exposed individuals as controls, the signal remained.

1204F
De novo deletion of 1q21.1 and Xq28 duplication in a family with developmental delay. K. Ha, H. Cho, L. Layman, H. Kim. Georgia Regents University, Augusta, GA.

Distal 1q21.1 microdeletion syndrome is a rare contiguous gene deletion disorder with de novo or autosomal dominant inheritance patterns. MECP2 duplication syndrome is an X-linked neurodevelopmental disorder characterized by intellectual disability, global developmental delay, and other neuropsychiatric complications including seizures later in life. Patients with these syndromes share some phenotypes such as mild dysmorphic features, cardiac abnormalities and mental retardation. Since these two syndromes are genetically unrelated, it has not been reported for these two different genetic syndromes to occur in the same family. Here we describe two siblings carrying a chromosome 1q21.1 microdeletion and chromosome Xq28 duplication. Using a comparative genomic hybridization (CGH) array, we identified a 1.24 Mb distal heterogeneous deletion at 1q21.1 from position 146,514,78 MB resulting in the loss of 12 genes and a 68 Kb deletion within intron 2 of AUTS2 gene at 7q11.22 in a girl with hypotrophydism, short stature, sensory integration disorder, and soft dysmorphic features including cupped ears and a unilateral ear pit. Protein kinase, AMP-activated beta-2 non-catalytic subunit (PRKAB2) regulating cellular energy metabolism, and Gap junction alpha-5 protein (GJA5), a gap channel protein, might have contributed the short stature, chronic constipation and sensory integration disorder in this patient. We also characterized a 508 Kb duplication at Xq28 encompassing MECP2 in the younger brother referred for hypotonia, poor speech, cognitive and motor impairment. In addition to MECP2, this duplication encompassed several other genes including L1CAM and FLNA that previously reported in other patients with a Xq28 duplication. The parental CGH analysis revealed that the 1q21.1 deletion in the elder sister is de novo, but the Xq28 duplication in the younger brother was inherited from the mother with a normal phenotype, suggesting a recessive pattern of genes involved in this region. We have described the candidate genes and phenotype-genotype relationships in two different chromosome regions of these two syndromes.
1205W
The first familial null mutation of an autism susceptibility gene in Autism Spectrum Disorder. N. Lambert1,2, V. Weremenko1, B. Pichon1, S. Acosta1, J. van den Ameelo3,4, C. Perazzolo5, D. Messina2, F. Musumeci2, B. Dessars3, A. De Leener1, M. Abramowicz4, C. Villain1. 1) ULB Center of Human Genetics, Hôpital Erasme, ULB, Brussels, Belgium; 2) Department of Paediatric Neurology, Hôpital Erasme, Université Libre de Bruxelles (ULB), Brussels, Belgium; 3) Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire (IRIHB), Université Libre de Bruxelles (ULB), Brussels, Belgium; 4) Department of Neurology, Ghent University Hospital, Ghent, Belgium.

The autism spectrum disorder results from interactions of genetic and environmental factors. Many genes have been described as candidate genes for autism susceptibility. The gene we studied is a candidate gene for autism susceptibility and is implicated in neurodevelopment and social brain circuit, but no mutation of this gene has ever been described. Here, we describe the first case of a familial mutation of this gene, consisting of an interstitial genomic deletion removing exons 12 through 15, causing a frameshift and premature stop codon, with evidence of nonsense-mediated mRNA decay. The heterozygous mutation was associated with autism in one patient and language and social impairment in a sibling. Our observations delineate the phenotypic spectrum associated with a clearly defined, very likely complete loss of function mutation of this gene. Incomplete penetrance in this family was consistent with this gene as a partial susceptibility gene for ASD. Implication of this gene in normal and pathological brain development opens new perspectives for understanding the pathophysiology of autism and for eventual therapeutic clues.

1206T
Use of a custom-designed array-CGH to identify small CNVs in nonsyndromic Autism Spectrum Disorders. C.M. Ribeiro1, E.S. Moreira1, A.L.B. Martins1, A.G. Morales1, S.G. Ferreira1, V.N. Takahashi1, D.P. Moreira1, K. Griesi-Oliveira1, A.C.F. Conte1, C. Rosenberg1, E. Vadassz1, M.R. Passos Bueno1. 1) Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, University of Sao Paulo, Sao Paulo, Brazil; 2) UNESP, Sao Jose do Rio Preto, Sao Paulo, Brazil; 3) Laboratorio de Genetica, Hospital de Base, Faculdade de Medicina de Sao Jose do Rio Preto, Sao Paulo, Brazil; 4) Instituto de Psiquiatria, Hospital das Clinicas, Faculdade de Medicina, University of Sao Paulo, Sao Paulo, Brazil.

Autism Spectrum Disorders (ASD) are highly heterogeneous and, despite the enormous efforts to identify their underlying genetic causes, at least 70% of the cases are idiopathic. We have hypothesized that a proportion of these cases might be caused by small CNVs not detected through conventional microarray platforms or NGS. To address this hypothesis, we designed an array-CGH with high density of probes targeted to the coding exons and to the UTR of 269 genes involved in biological processes potentially significant to the pathophysiology of ASD. A total of 508 ASD Brazilian patients and 550 Brazilian healthy individuals were included in the study after a signed informed consent. The IB-USP Ethics Committee approved this research. All patients were diagnosed by psychiatrists using DSM-IV-ICD-10 and ASQ, and all were negative for CNVs at 15q11-13, 16p11 and 22q13 through MLPA. Exclusion criteria comprised autism-related syndromes, metabolic disorders, facial dysmorphisms, congenital malformations, or prior exposure to known teratogenic drugs/ infectious agents during pregnancy. The number of individuals harboring at least one CNV and the number of CNVs per subject were significantly different between 103 cases and 200 controls (p<0.0001). These findings also applied to the separated analyses of the microdeletions and microduplications (p<0.0002), with odds ratios related to the number of alterations per individual between 2.92 and 3.32. Considering the subjects with two or more alterations, the differences between cases and controls were even greater (for microdeletions, p=0.00005; OR=28.74 [95% CI 3.7-223.1]). Forty-five out of 103 patients (43.7%) presented at least one CNV smaller than 1 kb which disrupted one of the genes in our customized array, a rate significantly higher than the proportion (21.5%) found in controls (p=0.00009). Finally, we detected 10 ASD-specific CNVs (five in the UTR) in 11.6% of the patients. These results are different from those usually obtained through conventional screenings with genome-wide platforms and indicate that the use of this high-resolution custom-designed array can increase the chance of detecting alterations enriched in ASD. This study also suggests that disruption of non-coding DNA may contribute to the risk factors for ASD, corroborates the association of SHANK2, DIAPH3, GCH1, GRM5 and MARK1 with nonsyndromic ASD, and reveals SNAP29, SLC17A6, PRKCa, MBD2 and GAD2 as new candidates for these disorders.

1207F
Linkage Disequilibrium Mapping of the 13q21 Specific Language Impairment Locus Using Epistasis Analysis Models. S.L. Wolock1, N. Li1, S.A. Petrelli2, J.F. Flax2, A.S. Basset2, L.M. Brzustowicz3, C.W. Bartlett1,2. 1) The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) Department of Psychology, The Ohio State University, Columbus, OH; 3) Department of Genetics, Rutgers University, Piscataway, NJ; 4) Department of Psychiatry, University of Toronto, Toronto, Canada; 5) Department of Pediatrics, The Ohio State University, Columbus, OH.

Background: We present a fine-mapping study of the 13q21 region linked to Specific Language Impairment (SLI), a neurodevelopmental failure to develop normal vocabulary and grammar despite otherwise normal cognition and ability. In previous work, we mapped the SLI3 locus to 13q21 using a written language impairment phenotype using five extended families from Canada. We replicated the locus using nuclear and extended families from the United States. In refining the localization, we found that a coding SNP in BDNF associated with memory greatly increased localization and evidence for the locus when included as part of a gene-gene interaction with the unidentified 13q21 risk alleles. Here, we performed additional mapping of the region to identify the gene responsible for the linkage signals. Methods: We assessed linkage/association of Illumina SNP array genotypes with eight phenotypes using the posterior probability of linkage (PPL) and the posterior probability of linkage disequilibrium (PPLD) metrics. We conducted two analyses: a baseline analysis and an analysis that incorporated BDNF genotypes from a coding SNP associated with memory into a gene-gene (epistatic) interaction model. The posterior probabilities from the two models are on the same scale and can be directly compared to assess the evidence that a SNP has an epistatic interaction. Results: Several SNPs within an LD block showed significant association with the quantitative trait underlying the written language impairment trait (maximum PPLD = 41%). We also observed weaker association with the quantitative trait and the categorical written language impairment diagnosis used to map this locus in the original genome scans (maximum PPLD = 41%). In both cases, modeling the coding SNP in BDNF showed an epistatic effect. The implicated region overlaps ATXN8OS, an anti-sense transcript of the KLHL1 gene. ATXN8OS contains a tri-nucleotide repeat that, when expanded, may increase the risk for a form of spinocerebellar ataxia. It has additionally been associated with the categorical written language impairment phenotype. Conclusion: While additional work to elucidate the role of this antisense transcript is ongoing, the findings suggest a role for BDNF-ATXN8OS interaction in the etiology of SLI.
Evidence for the involvement of MIR185 and its target genes in the development of schizophrenia. AJ. Forster1,2,9, FB. Basmanan1,2,9, M. Matthiesen1,4, A.C. Böhmer1,2, MV. Hollegaard3, E. Janson4, E. Strongman5, L. Priebe1,2, F. Degenhardt3,4, RA. Ophoff5,10,11, S. Moebus12, PB. Mortensen13,14, AD. Bargilum1,2,15, DM. Hougaard16, M. Rietschel14, A. Zimmer17, MM. Nöthen1,2, K. Miron1,2, K. Cichon1,2,17,19, 1) Institute of Human Genetics, University of Bonn, Germany; 2) Department of Genomics, Life and Brain Center, Bonn, Germany; 3) Department of Biomedicine, University of Aarhus, Denmark; 4) Department of Genomics Mathematics, University of Bonn, Germany; 5) Section of Neonatal Screening and Hormones, Statens Serum Institute, Copenhagen, Denmark; 6) Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands; 7) Division of Medical Genetics, University Hospital Basel and Department of Biomedicine, University of Basel, Switzerland; 8) Department of Psychiatry, University of Bonn, Germany; 9) Department of Psychiatry, University of Halle-Wittenberg, Halle, Germany; 10) Center for Neurobehavioral Genetics, University of California Los Angeles, USA; 11) Rudolf Magnus Institute of Neuroscience, Department of Psychiatry, University Medical Center, Utrecht, The Netherlands; 12) Institute of Medical Informatics, Biometry and Epidemiology, Essen, Germany; 13) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark; 14) Department of Biomedicine, University of Bonn, Germany; 15) Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark; 16) Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University, Germany; 17) Institute of Molecular Psychiatry, University of Bonn, Germany; 18) Institute of Neuroscience and Medicine, University of Bonn, Germany; 19) Medical Genetics, University Hospital, Utrecht, The Netherlands. The role of MIR185 in schizophrenia was investigated after correction for multiple testing. Resequencing identified two rare patient-specific novel variants directly flanking MIR185. However, follow-up genotyping provided no further evidence for their involvement in schizophrenia.

1209T

Idiopathic Basal Ganglia Calcification (IBGC) is characterized by bilateral calcification of the basal ganglia and presents clinically with a spectrum of neuropsychiatric and motor syndromes including dystonia, parkinsonism, tremor and chorea. Mutations in 2 genes, SLC20A2 and PDGFRB, have recently been shown to cause familial forms of IBGC. We set out to identify and study mutations in SLC20A2 and PDGFRB in a series of pathologically confirmed IBGC cases from the Mayo Clinic Florida Brain Bank. Methods: 27 cases were included based on the presence of significant basal ganglia calcification upon pathological examination. All coding regions of SLC20A2 and PDGFRB were sequenced. For SLC20A2, copy number analysis was also performed using real-time quantitative PCR. Confirmatory assays for a genomic deletion in SLC20A2 were performed by PCR based on the coordinates obtained from whole-genome sequencing (WGS) in one patient. For PDGFRB, we generated wild type and mutant forms and performed transient expression in HELa cells to evaluate its tyrosine kinase function. Upon stimulation with PDGF-BB, the autophosphorylation of wild-type and mutant PDGFRB at multiple tyrosine residues was assessed by western blot. Results: In our IBGC brain bank series, we identified two novel coding variants in SLC20A2. One novel variant was predicted to induce a nonsense mutation in exon 3 (p.S113X) and the second mutation consisted of a 563,256 bp genomic deletion with precise breakpoints defined by WGS. This deletion co-segregated with the disease in a large Canadian family and affected multiple genes. The combined deletion of SLC20A2 and THAP1 in this family likely contributed to the early onset and predominant dystonia. Additionally, FRB and FRB1 mutant PDGFRB mRNA and protein levels were assayed following stimulation with PDGF. Both FRB and FRB1 antisense constructs were found to reduce protein expression. Conclusions: Mutations in SLC20A2 were found in 3 cases of our IBGC series. The identification of a novel mutation and a complete deletion of SLC20A2 confirm an important role for SLC20A2 haplosufficiency in IBGC. Our work further supports a pathogenic role for PDGFRB in IBGC. Moreover, abnormal signaling through PDGFRB and deficient phosphate transport through SLC20A2 may be involved in the same pathogenic pathway leading to calcification in IBGC.
Extreme genetic heterogeneity among rare copy number variation in autism spectrum disorders.

M. Barbosa1, A.G. Chiocchetti2, D. Haslinger3, R. Wuttke1, E. Udekski2, T. Jarczok2, M. Sachsse2, F. Pousta2, A. Voran3, E. Huy3, F. Degenhardt4, S. Herms4,5, P. Hoffmann4,5, J. Meyer4,5, M.M. Biscaldi-Schäfer4, K. Schneider-Mommsen5, B.G. Schimmelmann5, S. Kupferschmid5, B. Herpertz-Dahlmann5, S. Gilsbach5, S. Moebus1,6, S. Cichon1,6, S.M. Klauck1, D. Pinto1, C.M. Freitag2, 1) Department of Psychiatry and Genetics and Genomic Sciences, Seaver Autism Center, Minich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, NY, USA; 2) Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe-Universität Frankfurt am Main, Address: Deutschorden-Str. 50, 60528, Frankfurt, Germany; 3) Department of Child and Adolescent Psychiatry and Psychotherapy, Saarland University, Homburg, Address: Klinikber-Str Building 90.2, 66421 Homburg / Saar, Germany; 4) Institute of Human Genetics, Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; 5) Division of Medical Genetics, University Hospital and Department of Biomedicine, University of Basel, Switzerland, Address: Mattenstr. 28, 4058 Basel, Switzerland; 6) Department of Neurobehavioral Genetics, University of Trier, Germany, Address: Johanniterufer 15, 54290 Trier, Germany; 7) Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital Freiburg, Address: Hauptstr. 6, 79104 Freiburg, Germany; 8) University Hospital of Child and Adolescent Psychiatry, Bern, Address: Bolligenstrasse 111, 3000 Bern, Switzerland; 9) Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital RWTH Aachen, Address: Neuenhofer Weg 21 in 52074 Aachen, Germany; 10) Hebrew University, Jerusalem, Israel; 11) Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany. Address: Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.

Autism spectrum disorders (ASD) is a group of complex neurodevelopmental abnormalities characterized by impairment in social interaction, language and communication deficits and repetitive behavior. The etiology for ASD has only been partially understood and the common etiology of this disorder is not yet known. Here, we report a pilot study to identify copy number variants (CNVs) with a high pathological potential for ASD and to clarify the complex genetic heterogeneity of the disorder in a well-characterized ASD cohort. A cohort of 716 German patients with ASD and their relatives (in a total of 2114 subjects) were assessed using the ADI-R and ADOS diagnostic tools. The control group is composed of 1320 German subjects ascertained through the Heinz Nixdorf Recall Study. Both cohorts have been genotyped on the Illumina Human OmniExpress 730K-SNP-array. For copy number variation (CNVs) analysis, we used 4 calling algorithms: PennCNV, QuantiSNP, GNOSIS, CNVPartition. A total of 1967 ASD-samples passed stringent quality control, 673 being probands (572 males and 101 females). Of these, 394 were from trios, 42 from quads and 20 from large families. Using also data from an additional 5000 European controls, priority was given for rare CNVs (de novo or inherited).

A variety of _de novo_ CNVs was detected in cases, from large aberrations (e.g. a novel 13q14.2 5Mb-deletion) to rare pathogenic CNVs (e.g. a Potocki-Lupski 17q12-duplication-syndrome). Interestingly, inherited CNVs along with second genomic hits were seen in four subjects (e.g. a maternal NRXN1 intragenic deletion and _de novo_ 22q11.2-duplication), giving support to a multiple-hit hypothesis. Also, we found CNVs that are often related to a remarkably variable neurobehavioral phenotype such as the 3q29-deletion syndrome, which can present as intellectual disability, ASD, psychosis, anxiety, hyperactivity and/or aggressiveness. Further experimental validation and network modelling of genes intersected or disrupted by CNVs are ongoing.
1212T Transcriptome outlier analysis implicates schizophrenia candidate genes as harboring rare functional variants of large effects. J. Duan1, W. May1, J. Freda1, E.I. Drigalenko1, H.H.H. Göring1, A.R. Sanders1, P.V. Geijman1. Molecular Genetics of Schizophrenia (MGS) Collaboration. 1) Dept Psychiatry, Northshore Univ Healthsystem/Univ of Chicago, Evanston, IL; 2) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.

Multiple rare and large copy number variants (CNVs) are associated with high risk for schizophrenia (SZ), and some common risk variants have been uncovered by GWAS. However, a substantial part of the genetic risk still remains unexplained and disease mechanisms are largely unknown. To uncover additional loci of large effect, we propose to sequence targets derived from transcriptome analyses. Most gene expression studies rely on the study of average abundances in case-control samples. Although useful, this approach assumes substantial euticogenic homogeneity, and may miss rarer genetic effects. We have employed an alternative approach to identify expression outliers (i.e., expression distribution extremes) where the tails of the distribution are enriched for cases. We hypothesize that such case-enriched expression outliers are caused by rare coding variants (e.g., nonsense or splice site mutations) or regulatory mutations of large effect on gene expression. We analyzed an RNAseq dataset comprised of lymphoblastoid cell lines from 312 SZ cases and 322 controls, none of which carry a known SZ-associated CNV. We calculated the Z-scores of the expression for each gene for the whole sample. Expression outliers were defined as genes with abundances beyond a predefined standard deviation cutoff (2SD or 3SD). We identified 828 expression outlier genes with 2SD-tails (401 lower and 427 upper) enriched for SZ cases. In these genes we observed enrichment of brain-expressed genes, SZ-risk CNV-spanned genes, and genes within CNVs associated with neurodevelopmental disorders. Note-worthy outlier genes: KCTD13 that drives the 16p11.2 CNV phenotype, DSCR8 at 22q11.21 deletion region, and TCFC3 belonging to the same gene family as the SZ-associated TCF4. We did not find global outlier burden differences in SZ cases and controls (i.e., total number of outlier genes, individual). However, SZ cases showed higher outlier burden for genes spanned by the above SZ-risk CNVs and a trend towards higher outlier burden for genes with abundances beyond 3SD. These results suggest that genes identified by outlier expression analysis are relevant to SZ pathogenesis. However, given the moderate sample size and multiple statistical tests, we consider our results exploratory. We are currently sequencing these outlier genes and will report the functional variants contributing to aberrant mRNA expression in SZ cases.

1213F Molecular genetic study on a Japanese family with amyotrophic lateral sclerosis. Y. Ichikawa1, H. Ishiura1, J. Mitsu1, S. Endo1, Y. Takahashii1, A. Ishi1, M. Watabas1, A. Tamaoka1, M. Murata1, K. Doi1, J. Yoshimura1, S. Morishita1, J. Goto1, S. Tsur1. 1) Dept Neurology, Univ of Tokyo, Tokyo, Japan; 2) Dept Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan; 3) Dept Neurology, Univ of Tsukuba, Ibaraki, Japan; 4) Dept Computational Biology, Univ of Tokyo, Tokyo, Japan.

Background: Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder characterized by degeneration of motor neurons in the brain and spinal cord. Although the majority of ALS patients are sporadic, approximately 10% of ALS cases are familial (FALS). In the Japanese population, mutations of SOD1, FUS, TARDBP, and C9orf72 account for approximately 60% of the cases of FALS, and the causative genes remain to be elucidated in 40% of the cases of FALS. Objective: This study is aimed to search for a causative gene for a Japanese autosomal dominant FALS family. Subjects and Methods: Three patients were affected in this family. The age of onset of one patient was 12 years, other two patients were in their 70s. Concerning the clinical presentations, one was progressive bulbar palsy; the others were Aran-Duchenne type. Genomic DNAs were extracted from nine individuals including two affected patients. Genotyping was carried out using Affymetrix SNP Array 6.0 and analysis was performed using allelo and SNP HitLink. To explore the causative mutations, we conducted an exome analysis of the index patient. Results: The linkage analysis revealed the highest multipoint parametric LOD score of 1.8 in 13 regions on chromosomes 1, 4, 5, 6, 8, 9, 10, 11, 12, and 14 spanning 189.5Mb in total. Although, the loci of the causative genes for dominant FALS including 11, 12, and 14 spanning 189.5Mb in total. Although, the loci of the causative genes for dominant FALS including SETX, ANG, and TARDBP were included in the candidate region, the mutations of these genes were not detected by the exome analysis. Conclusion: The results of the analyses suggest that ALS in the present family is caused by a mutation of a novel gene.

1214W Association between serotonergic transporter variant, 5-HTTLPR, and hoarding traits in a population-based pediatric sample. V. Sinopoli1, A. Dupuis2, C. Burton2, J. Crosby2, J. Shan1, R. Schachar1,3,6, J. Tsai1,2,6, M. Potlord1,2,6, M. Gejman1, J. Duan1, J. Driesen1,1. 1) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 2) The Hospital for Sick Children, Toronto, ON, Canada.

Obssesive-compulsive disorder (OCD) is an anxiety disorder characterized by recurring, intrusive thoughts or compulsive behaviors aimed at alleviating the resulting anxiety. It is a heterogeneous disorder that is both genetically and phenotypically complex, which complicates our ability to identify the heritable components of OCD and the genes that are responsible. At the genetic level, studies reveal several promising lines of evidence within the disorder. At the genetic level, the promoter region of the serotonin transporter gene, 5-HTTLPR, has been implicated in OCD, but findings have been inconsistent. Statement of Purpose: The aim of this study is to examine a functionally triallelic variant in 5-HTTLPR and its genetic association with obsessive-compulsive (OC) traits, as a whole, and specific subgroups of OC traits in a population-based sample of children. Methods: We have identified OC traits in 7,500 Caucasian children, from the Ontario Science Centre in Toronto, and have selected the top and bottom 10% of our sample for follow-up. Following collection and extraction of DNA, we directly genotyped the 5-HTTLPR polymorphism and classified variants based on functionality. We examined the association between 1) genetic presentation of the variant and high versus low OC traits and 2) genetic presentation of the variant and high versus low OC traits within each factor-based subgroup we have identified. No significant difference is seen between the high OC trait group and the low OC trait group, as a whole. A significant difference in allelic and genotypic distribution is seen between individuals with and without OC traits. Our findings highlight the importance of functional factors. The LA/LA genotype occurs more often with low extreme hoarding traits (P-value 0.004). Likewise, the LA allele occurs more often with low extreme hoarding traits, compared to any other allele (P-value 0.003). Conclusion: Our findings reveal a significant difference in genotypic and allelic distribution only between individuals with versus low extreme hoarding scores. This illustrates a need to consider specific symptom dimensions within OCD, when examining serotonergic and other genetic variants. By reducing pheno-type heterogeneity, our approach promises to more accurately facilitate identification of genetic risk factors in OCD.

1215T Disruption of a large intergenic noncoding RNA, FLJ42707, is associated with developmental delay and language disorder. S. Fan1, L. Dukes-Rimsky1, S. Ladd1, B.R. DuPont3, C. Skinner1, L. Wang4, K.B. Clarkson1, A.K. Sivivastava1,2, J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC; 2) Department of Genetics and Biochemistry, Clemson University, Clemson, SC.

Large intergenic noncoding (lnc) RNAs are emerging as key regulators of diverse cellular processes including neurodevelopment. However, the function of the majority of these lincRNAs remains uncharacterized. We report a novel intergenic, balanced chromosomal breakpoint at 1q23.1 and 15q15 in a 2 year and 7-month-old male patient with developmental delay, speech delay, significant expressive language delay, and congenital infantile left eye esotropia. Fluorescence in situ hybridization and whole genome sequencing confirmed that the chromosomal breakpoints and revealed the disruption of a previously uncharacterized lincRNA gene, FLJ42707 at 5q15. The breakpoint at 1q23.1 disrupted no gene. Further studies by PCR amplification and sequencing of the breakpoint junctions revealed the identical breakpoints in the parents of the patient, as well as in two other patients. We sequenced the breakpoint junctions in 100 controls but only in one case did we observe the lincRNA expression of several genes flanking each chromosomal breakpoint and found no change in their expression levels than control individuals. In addition, we looked for disease causing variants other than the transcription breaks using the InGeny Variant Analysis tool (InGeny Systems Inc, CA). No potential disease causing variant was found. Thus, we further characterized FLJ42707 and confirmed expression of at least three major transcribed isoforms of FLJ42707. We further determined that all three isoforms were expressed at a higher level than in controls and in most other individuals tested. To infer the physiological function of FLJ42707, we looked for genes that are co-expressed with FLJ42707. On the basis of probe-to-probe correlation coefficient calculated from an integrated set of 2.968 microarray expression profiles of healthy human tissue samples, we found that FLJ42707 was co-expressed with the expression of FLJ42707 and with genes enriched with Gene Ontology terms including synaptic transmission, cell-cell signaling, neuron projection, and synapse. Recently, several genomic deletions at 5q15 encompassing FLJ42707, NR2F1, FAM172A, PU59SF2, and MIR227 gene regions have been reported in patients with velocardiofacial syndrome, strabismus/esotropia, deafness and significant delay in gross motor skills, cognitive and communication skills. All together, our findings raised the possibility of a likely role for the lincRNA gene FLJ42707 in neurodevelopmental disorders.
Gene-based association analysis of Alzheimer’s disease risk gene CD33 with brain amyloid-β burden and microglial activity measured by PET. K.T. Nho1,2, S. Shanker1, S. Khachaturian1,2,3, L. S. Risacher4, V.K. Ramanan1,2, L. Shen1,2, K.K. Yoder1, P.S. Aisen3, R.C. Petersen4, M.W. Weiner5,6,7, B.C. McDonald1,3,8, E.F. Tallman1,3, G.D. Hutchins1,3, J.W. Fletcher1, M.R. Farlow9,8, B. Ghetiri10, A.J. Saykin1,2,3,8 for the Alzheimer’s Disease Neuroimaging Initiative (ADNI). 1 Radiology and Imaging Sciences, Indiana University, Indianapolis, IN; 2 Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN; 3 Indiana Alzheimer Disease Center, Indianapolis, IN; 4 Department of Neurosciences, University of California, San Diego, San Diego, CA; 5 Mayo Clinic, Rochester, MN; 6 Departments of Radiology, Medicine and Psychiatry, University of California, San Francisco, San Francisco, CA; 7 Department of Veterans Affairs Medical Center, San Francisco, CA; 8 Department of Neurology, Indiana University School of Medicine, Indianapolis, IN; 9 Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN; 10 Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN. Background: Large-scale GWAS have identified and confirmed CD33 significantly associated with late onset Alzheimer’s disease (LOAD) (Naj et al., 2011). A recent study showed CD33 is expressed in microglial cells in the human brain and inhibits microglial uptake of amyloid-β (Griciuc et al., 2013). Our aim was to perform gene-based association analyses to investigate the relationship of CD33 variants with amyloid-β (Aβ) burden measured using florbetapir positron emission tomography (PET) imaging (Ramanan et al., 2013) (Analysis 1) and microglial activity measured using [11C]PBR28 PET imaging (Yoder et al., in press) (Analysis 2). Methods: 6,720 participants in the ADNI (Alzheimer’s Disease Neuroimaging Initiative) cohort with florbetapir PET scans were used in Analysis 1, and 26 participants in the IMAG (Indiana Memory and Aging Study) cohort with [11C]PBR28 PET scans were used in Analysis 2. The endophenotypes used in Analysis 1 were regional standard uptake value ratio (SUVR) values from 5 regions adjusted for age, gender, diagnosis, APOE ε4 status, and BCHE genotypes (Ramanan, et al. 2013) and in Analysis 2 were mean PBR SUV from 8 regions adjusted for T3PPO genotype. SNPs within ±10 bp of the 3’ end of the gene was identified with the minor allele associated with PBR SUV (p<0.072) in many regions. In Analysis 1, rs12971800 located at 4kb of the CD33 gene after imputation using MACH were included. Gene-based association analysis and independent if rs12971800 was significantly associated with LOAD (Naj et al., 2013). Our aim was to perform gene-based association analyses to perform gene-based association analyses to perform gene-based association analyses to perform gene-based association analyses to perform gene-based association analyses to perform gene-based association analyses to perform gene-based association analyses to perform gene-based association analyses to perform gene-based association analysis of Alzheimer’s disease risk gene CD33 with brain amyloid-β burden and microglial activity measured by PET. K.T. Nho1,2, S. Shanker1, S. Khachaturian1,2,3, L. S. Risacher4, V.K. Ramanan1,2, L. Shen1,2, K.K. Yoder1, P.S. Aisen3, R.C. Petersen4, M.W. Weiner5,6,7, B.C. McDonald1,3,8, E.F. Tallman1,3, G.D. Hutchins1,3, J.W. Fletcher1, M.R. Farlow9,8, B. Ghetiri10, A.J. Saykin1,2,3,8 for the Alzheimer’s Disease Neuroimaging Initiative (ADNI). 1 Radiology and Imaging Sciences, Indiana University, Indianapolis, IN; 2 Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN; 3 Indiana Alzheimer Disease Center, Indianapolis, IN; 4 Department of Neurosciences, University of California, San Diego, San Diego, CA; 5 Mayo Clinic, Rochester, MN; 6 Departments of Radiology, Medicine and Psychiatry, University of California, San Francisco, San Francisco, CA; 7 Department of Veterans Affairs Medical Center, San Francisco, CA; 8 Department of Neurology, Indiana University School of Medicine, Indianapolis, IN; 9 Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN; 10 Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN. Background: Large-scale GWAS have identified and confirmed CD33 significantly associated with late onset Alzheimer’s disease (LOAD) (Naj et al., 2011). A recent study showed CD33 is expressed in microglial cells in the human brain and inhibits microglial uptake of amyloid-β (Griciuc et al., 2013). Our aim was to perform gene-based association analyses to investigate the relationship of CD33 variants with amyloid-β (Aβ) burden measured using florbetapir positron emission tomography (PET) imaging (Ramanan et al., 2013) (Analysis 1) and microglial activity measured using [11C]PBR28 PET imaging (Yoder et al., in press) (Analysis 2). Methods: 6,720 participants in the ADNI (Alzheimer’s Disease Neuroimaging Initiative) cohort with florbetapir PET scans were used in Analysis 1, and 26 participants in the IMAG (Indiana Memory and Aging Study) cohort with [11C]PBR28 PET scans were used in Analysis 2. The endophenotypes used in Analysis 1 were regional standard uptake value ratio (SUVR) values from 5 regions adjusted for age, gender, diagnosis, APOE ε4 status, and BCHE genotypes (Ramanan, et al. 2013) and in Analysis 2 were mean PBR SUV from 8 regions adjusted for T3PPO genotype. SNPs within ±10 bp of the 3’ end of the gene was identified with the minor allele associated with PBR SUV (p<0.072) in many regions. In Analysis 1, rs12971800 located at 4kb of the CD33 gene after imputation using MACH were included. Gene-based association analyses were performed using a dominant genetic model and SNPs were considered significant if they had a p<0.05 on linear regression analyses and independent if r²<0.05. Permutation was used to correct for multiple SNPs. Results: The Gene-based association analysis revealed that the CD33 gene was significantly associated with Aβ SUVR (p<0.0044) and marginally PBR SUV (p<0.072) in many regions. In Analysis 1, rs12971800 located at the 3’ end of the gene was identified with the minor allele associated with higher Aβ SUVR and accounted for ~2% of the phenotypic variation. In Analysis 2, another SNP (rs273645) located at the 3’ end of the gene was identified with the minor allele associated with lower PBR SUV. Conclusions: CD33 was associated with PET imaging measures associated with AD, suggesting CD33 appears to contribute to known neuropathologic features of AD such as amyloid burden and inflammation. This is relevant in light of the previous findings and warrants further investigation in independent and larger samples.

Secretin is a peptide hormone released from duodenum to stimulate exocrine secretion from pancreas. It is also produced in the brain and functions as a neuropeptide hormone in the central nervous system. Secretin receptor (SCTR) is a member of the type G-protein coupled receptor family, and expressed in the brain. In the rat brain during the early postnatal period, an active synaptogenesis period, expression levels of secretin and SCTR mRNA are higher than in the adult brain. (Tay, et al. Neuroscience letters. 2004) Since SCTR-deficient mice show abnormal social and cognitive behaviors, furthermore, synaptic plasticity in the hippocampus is impaired (Nishijima, et al.HMG.2006), the involvement of secretin to the pathogenesis of autism spectrum disorder (ASD) is suggested. To investigate the contribution of secretin and its pathway to ASD, we screened the SCTR gene for mutations in ASD patients. (Subjects and methods) We analyzed 197 DNA samples from Japanese ASD patients with the parent’s informed consent and Caucasian patients from the Autism Genetic Resource Exchange ( AGRE). Each exon and adjacent intron of SCTR was amplified by PCR and subjected to direct sequencing. Base changes detected in ASD samples were screened in 126 Japanese and Caucasian control samples. (Results) We detected three missense mutations in the SCTR gene that were not detected in controls, the R2C in one Caucasian male patient, the P90L in one Japanese male patient and the A245T in one Caucasian male patient. We also found another base substitution in the SCTR gene, the F411S in one Japanese male patient, but also in one control. (Discussion) These mutations are conserved regions. The R2C mutation in the SCTR is located in signal peptide domain and the P90L mutation is located in hormone binding domain. These mutations can have effect to alter the secretin binding to the receptor and its function. These findings suggested that secretin pathway play an important role in the pathogenesis of ASD.


Phospholipid Flippase ATP8A2 is essential for Normal Visual and Auditory Function and Photoreceptor and Spiral Ganglion Cell Survival. X. Zhu, Y. Zhou, X. Zhu, 1) The Sichuan Provincial Key Laboratory for Human Disease Gene Study and The Institute of Laboratory Medicine, Sichuan Academy of Medical Sciences & Sichuan Provincial People’s Hospital, Chengdu, Sichuan, China; 2) Sichuan Translational Medicine Hospital, Chinese Academy of Sciences, Chengdu, Sichuan, China.

ATP8A2 is an important P4-ATPase phosphatidylserine flippase, which is required for proper maintenance of neuron and axon function. ATP8A2 is highly expressed in the brain, spinal cord, testes and retina. In the retina, ATP8A2 is primarily expressed in the light-sensitive photoreceptor cells and localized in the outer segment compartment. Although mutations in ATP8A2 have been reported to cause mental retardation in humans and spinal motor neuronal degeneration in the wobbler-lethal (wl) mouse, the physiological and cellular role of ATP8A2 in sensory systems have not been investigated. In this study we have analyzed the retina and cochlea of Atp8a2 mutant mice in order to determine the role of ATP8A2 in visual and auditory systems. ATP8A2-deficient mice had shortened photoreceptor outer segments, a significant reduction in rod and cone photoreponses, and decreased photoreceptor viability. The ultrastructural organization of photoreceptor outer segments appeared normal. The auditory brainstem response threshold was significantly higher in the ATP8A2-deficient mouse. The inner and outer hair cells appeared normal, but a significant degeneration of spiral ganglion cells was apparent. Our studies indicate that ATP8A2 plays a crucial role in photoreceptor and spiral ganglion cell function and survival by maintaining phospholipid composition and asymmetry and contributing to vesicle trafficking in these neurons.

Application of causal inference methods to investigate APOE variation and cognitive impairment in multiple sclerosis. E. Elkind-Wileray, P. Yousef, L. Shen, M.F. George, H. Quach, F.B.S. Briggs, A. Bernstein, C. Schaefer, L.F. Barcellos. 1) Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, UC Berkeley, Berkeley, CA; 2) Division of Environmental Health Sciences, School of Public Health, UC Berkeley, Berkeley, CA; 3) Division of Research, Kaiser Permanente, Oakland, CA; 4) Palm Drive Hospital, Petaluma, CA, USA.

1223W
Social skills impairments in girls with Turner syndrome. M. Inbar-Feigenberg1,2, D. Grafodatskaya3, S. Choufani4, BHY. Chung1,2,3, L.J. Roberts1, C. Russell5, W. Roberts1, J. Hamilton6,6, R. Weksberg1,2,6 1) Clinical & Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Genetics and Genome Biology Program, Hospital for Sick Children, Toronto, ON, Canada; 3) Centre of Reproduction, Growth & Development, Department of Pediatrics & Adolescent Medicine, The University of Hong Kong, Hong Kong; 4) Autism Research Unit, Hospital for Sick Children, Toronto, ON, Canada; 5) Division of Endocrinology, Department of Pediatrics, Hospital for Sick Children, Toronto, ON, Canada; 6) Department of Pediatrics, University of Toronto, Toronto, ON, Canada.

Introduction: Turner syndrome (TS) is one of the most common sex chromosome abnormalities caused by complete or partial monosomy of the X chromosome, with a prevalence of ~1/2000 female live births. Individuals with TS often exhibit with short stature, gonadal dysfunction and other systemic malformations. A specific neuro-cognitive profile has been reported, sometimes including impaired social cognition. Autism and autism spectrum disorders (ASD) are reported in 5% and 25% of TS patients respectively. Skuse et al. 1997 reported that females who inherit their single X chromosome from their father have better social skills than females who inherit it from their mother. The authors hypothesized that an imprinted locus on the X chromosome is relevant to social functioning. Hypothesis: Females with TS demonstrate parent of origin-specific differences in social cognition. Methods and Results: We recruited 28 individuals with TS (age 3–18 years) and their parents at the Pediatric Endocrinology Clinic. We collected buccal samples from the proband and both parents. In addition, parents completed two social skills questionnaires for their daughters, one used originally by Skuse et al. and the Social Responsiveness Scale (SRS) that assesses social awareness, social cognition, social communication, social motivation, and autistic mannerisms. SRS total scores fall into three categories: normal (T score ≤59); mild to moderate, consistent with mild ASD (T score=60–75); and severe, consistent with autism (≥76). t-tests were used to compare total scores for groups of girls with TS carrying a single maternal vs paternal X chromosome, as well as a group with karyotypes other than XO. In 14/28 patients (50%) scores were >60. In 7/28 (25%), scores were in the mild/moderate ASD range and in 7/28 (25%), scores were in the severe autism range. Score differences for the sub-scales of SRS showed higher scores for autistic mannerisms. A good correlation between Skuse et al. and the currently SRS was found (R²=0.80). We did not find any correlation in social skill measures with TS pregnancies of origin of the X chromosome in our small TS cohort. Conclusions: We found the rate of autistic features in TS to be significantly higher than previously reported. These data have significant implications for genetic counseling. We suggest that individuals with TS be routinely screened for ASD for early identification and initiation of behavioral interventions.

1224T
Examining the role of the 1q deletion critical region gene SCCPDH in early brain development in zebrafish. E.A. Burke1,2, W.A. Gahl1, C.F. Boeker1,3. 1) NIH Undiagnosed Diseases Program Translational Laboratory, Bethesda, MD; 2) Section on Human Genetic Genetic, Medical Genetics Branch, NHGRI, NIH, Bethesda, MD.

Submicronic terminal 1q deletion results in a syndrome consisting of severe psychomotor delays, aphasia, hypotonia, microcephaly, corpus callosum abnormalities, and facial dysmorphism. Though previous studies have narrowed the critical region for this phenotype to 1q43-1q44, the deleted genes responsible for the syndrome remain unknown. We hypothesized that a potential contributor to this syndrome is SCCPDH, which lies within the 1q44 critical region and is expressed in the developing human and zebrafish brain. Knockdown of the zebrafish homologue using morpholinos caused a dose-dependent loss of brain tissue at the midbrain-hindbrain boundary, decreased eye size, and hydrocephalus. Early phenotypic rescue was by co-injection of wild type SCCPDH mRNA in over 80 percent of embryos. Overexpression of wild type mRNA in zebrafish resulted in an increased incidence of cyclopia and abnormal forebrain development. In situ hybridization, SCCPDH morphants were found to widely overexpressed in the optic primordia in a manner similar to embryos injected with shh (sonic hedgehog) mRNA. These results, along with gene expression analyses, suggest a potential link between SCCPDH and Shh-related signaling. Therefore, we conclude that SCCPDH expression is highly regulated during brain development and that one mechanism by which it may affect brain development is through modulation of Shh signaling.

1225F
Investigation of the role of serotonin pathway in the etiology of Tourette’s syndrome. N. Sun1,2, Z. Pang1, L. Deng1,2, R.A. King1,2, M. Shelton1,2,3, J.C. Moore1,3, R.P. Hart1,4, M. Konstoksil1,2, G.A. Heiman1,2, J.A. Tischfield1,2,3. 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Human Genetics Institute of New Jersey, Piscataway, NJ; 3) RUCDR Infinite Biosciences NiMH Stem Cell Resource, Piscataway, NJ; 4) Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ; 5) Child Health Institute of New Jersey, New Brunswick, NJ; 6) Yale Child Study Center, Yale University School of Medicine, New Haven, CT.

Tourette’s syndrome (TS) is a childhood onset neurodevelopmental and genetic disorder characterized by chronic motor and phonic tics. Previous studies suggest the etiology of TS is the result of interactions between multiple genetic and environmental factors, but specific genetic factors contributing to TS causation are poorly understood. A recent publication documented that an I425V mutation in the serotonin transporter gene (SLC6A4) is found in certain subsets of TS patients, suggesting the involvement of the serotonin pathway in the etiology of TS. It is unclear how this mutation leads to altered expression of genes in neurotransmitter pathway(s). We collected blood samples from all five family members of a two-generation pedigree from the New Jersey Center for Tourette Syndrome Sharing Repository. Four individuals from this family, diagnosed with TS or/and Obsessive-Compulsive Disorder (OCD), carry the I425V mutation. We established induced pluripotent stem cell (iPSC) lines from each individual in this family. In order to develop a neuron-based TS model in vitro, we induced iPSC lines into neural stem cells (NSCs) and further differentiated the NSCs into electrophysiologically functional neurons. By performing whole cell patch clamp analysis on these neurons, we are assessing whether or not the I425V mutation leads to electrophysiological changes. In addition, we are conducting gene expression analyses of NSCs and neurons to identify any differentially expressed genes in neurotransmitter pathways involving serotonin which may lead to the discovery of epigenetically-modified genes involved in disease causation and development.

1226W
Distinguishing Autism Spectrum Disorders From Other Developmental Delays Using Blood RNASeq. S. LETOVSKY1,2, M. Causey1, M. Ayre2, J. Sokolstyk1, C. Proulx1, F.R. Sharp1, I.N. Pessah1, R. Hansen1, J. Gregg2, I. Hertz-Picciotto2, 1) SYNAPOD, Woburn, MA; 2) Massachusetts General Hospital, Boston, MA; 3) MIND Institute, University of California, Davis, CA.

There is an unmet need for objective biomarkers to assist clinicians in the early diagnosis of childhood neurodevelopmental disorders. The aim of this study was to assess whether blood gene expression measured using next generation RNA sequencing (RNASeq) could provide a biomarker to distinguish children on the autism spectrum from children with other conditions that might present in the same clinical setting. The CHARGE (Childhood Autism Risks from Genetics and the Environment) study recruited children between the ages of 2 and 5, some of whom were diagnosed on the autism spectrum, and others with other developmental delays. Subjects were grouped based on thresholds of the ADOS, ADI-R, Vineland and Mullens test into autism spectrum disorder (ASD) and other developmental delay (DD) groups to approximate the clinical use case of a secondary screen for autism in children suspected of neurodevelopmental disorders. RNASeq was performed on RNA from blood tubes in 174 ASD and 96 DD samples passed final QC, for a total of 270 samples. Sequence data were processed through the Tuxedo RNASeq pipeline. Samples were divided into a training set (n=153) and a holdout set (N=117), each of which was repeatedly subsampled to achieve gender and age balance between the ASD and DD groups. On each iteration, informative features were selected by t-test and a support vector machine classifier was trained on a balanced subsample of the training set and tested on a balanced subsample of the holdout set. AU-C’s (area under the ROC curve) were averaged across iterations. The mean AUC for the holdout set was 65.6 +/- 2.9%. When a 90% sensitivity threshold was selected on the classifier risk score, the mean specificity was 25.3, with 95% CI [13.6, 40.6%]. Gene categories found significant by ranksum test on the t-statistic included immune response, cell cycle, immune and inflammation-related GO categories. To our knowledge this represents the first report of a classification signature for ASD vs. DD using blood RNASeq. A gene expression signature with moderate AUC has potential clinical utility as a sensitive assay for distinguishing autism from other developmental disorders. Planned followup studies include a multicenter clinical study to further refine and validate a blood-based assay.
Mutation spectrum in the dystrophin gene disclosed by MLPA in 181 Vietnamese Duchenne/Becker muscular dystrophy patients. K.V. Tran1, H.N. Do1, T.H. Tran1, M.H. Ta1, A.T.P. Le1, D.C. Vu1, V.T. Ta1, M. Matsuo2. 1) Center for Gene and Protein Research, Hanoi Medical University, Hanoi, Viet Nam; 2) Department of Medical Genetics, Metabolism & Endocrinology, National Hospital of Pediatrics, Hanoi, Vietnam; 3) Department of Medical Rehabilitation, Faculty of Rehabilitation, Kobegakuin University, Kobe, Japan.

Duchenne/Becker muscular dystrophy (DMD/BMD) is the most common X-linked muscular dystrophy caused by mutation in dystrophin gene. Deletion and duplication in the dystrophin gene account for 60-70% of mutation. Multiplex ligation-dependent probe amplification (MLPA) is the most powerful and convenient method to identify exon deletions or duplications in the dystrophin gene because of its overall gene coverage. The present investigation was designed to detect mutation in the dystrophin gene in 181 unrelated Vietnamese Becker/Duchenne patients using MLPA analysis. Among the 181 cases, deletions and duplications encompassing one or more exons were identified in 105 (58%) or 12 (6.6%) cases, respectively. Deletions were found to cluster in the proximal (14.3%) and central hotspot regions (72.4%); 14% were observed to have gross deletions and 1.2% had deletion including NXPH2 gene. All of those regions include autism, intellectual disability or schizophrenia. Previously, only 3 cases of deletion including NXPH2 were reported. These results indicated that intellectual disability of this patient results from haploinsufficiency of NXPH2.

Autism spectrum disorders (ASDs) are common neurodevelopmental conditions whose biological basis has been largely undetermined. Although the ASD prevalence has risen to 2% of the school-aged US population and the conditions have a tremendous impact on society and families, the current diagnosis is solely based on the analysis of the complex behavioral phenotype and is therefore often delayed. We have recorded the metabolic profile of lymphoblastoid cell lines from 87 patients with ASDs, 50 with non-ASD neurodevelopmental disorders, and 78 normal individuals, using Biolog Phenotype MicroArrays (PMs). This assay is designed to evaluate the production of energy (NAH) in cells via diverse metabolic pathways. The ASD cohort showed reduced energy production when tryptophan was the sole energy source. The abnormality in tryptophan metabolism correlated with both syndromal and non-syndromal ASD, independent of the genetic background of the individual. This metabolic alteration was not observed in a previous gene expression study of 10 patients with ASDs found abnormal levels for some genes involved in tryptophan metabolic pathways. Preliminary results indicate that the decreased tryptophan metabolism is detectable even in fresh blood leukocytes from patients with ASDs, suggesting that the assay might be further simplified and optimized into a quick and reliable screening test applicable at an early age and a precursor of serotonin, kynurenic acid, and kynurenine acid which are involved in neurodevelopment and synaptogenesis. Quinolinic acid is the structural precursor of NAD+, a critical energy carrier in mitochondrial and the serotonin branch of the tryptophan metabolic pathway generates NADH. Additionally, the levels of quinolinic and kynurenic acid are strongly influenced by the activity of the immune system. Therefore, decreased tryptophan metabolism may alter brain development, neurotransmitter activity, and mitochondrial function. Our finding of decreased tryptophan metabolism in primary cilium from patients with ASDs, suggests that the assay might be further simplified and optimized into a quick and reliable test applicable at an early age.

1232W Genetic evaluation of hereditary spastic paraplegia. S. Schirmer1, Z. Kohli2, T. Rödl1, M. Bocher1, J. Winkler2, U. Hehr1,3. 1) Center for Human Genetics, Regensburg, Germany; 2) Division of Molecular Neurology, University Hospital Erlangen, Erlangen, Germany; 3) Department of Human Genetics, University of Regensburg, Regensburg, Germany.

Hereditary spastic paraplegias (HSPs) comprise a clinically and genetically heterogeneous group of neurodegenerative disorders with progressive degeneration of the corticospinal tract. Patients with complicated forms show additional clinical features e.g. thin corpus callosum or peripheral neuropathy overlapping with a wide spectrum of other underlying genetic conditions. Currently 56 HSP loci have been assigned with causal mutations identified in 37 genes. Genetic testing by conventional Sanger sequencing in the diagnostic setting is usually restricted to the genes, most frequently affected. We here report the results of genetic testing for HSP by linkage analysis, Sanger sequencing and MLPA over the last 10 years. A subgroup of 20 patients received extensive neurologic workup in our outpatient clinic for movement disorders (cohort 1). In the overall cohort of 176 patients with suspected pure and complicated HSP causal mutations were identified in SPG4 (n=39), SPG3a (n=8), SPG7 (n=2), SPG11 (n=2), SPG20 (n=2) and SPG31 (n=2). To further increase the diagnostic yield we now evaluated a virtual gene panel and extended sequencing of currently 30 HSP genes (step 1) based on the Illumina TruSight™Exome assay, which for further diagnostic workup of mutation negative patients allows subsequent evaluation of additional candidate genes (step 2) associated with other movement disorders. Diagnostic yield was increased to 56% by further evaluating an additional 31 genes (cohort 2). In the overall cohort of 128 HSP patients with known mutations. Sample preparation was processed with the TruSight™Exome Enrichment kit. The Library was sequenced as a 150bp paired-end run with the MiSeq Reagent Kit v2. Reads were aligned to the human genome reference (UCSC hg19, NCBI build 37) and evaluated with an in house workup including the DNASTAR Lasergene®SeqMan Pro™ and DNAStar ArrayStar® software. With our step 1 protocol we reached an average coverage of 58x with a minimal coverage of more than 20x for 91% of all 460 analyzed targets. All 128 patients were fully analyzed and mutations were reliably detected including nonsense (n=2), missense (n=7) and frameshift (n=1) mutations, one 5bp deletion and one 9bp duplication. Clinically selected families of cohort 1 with suspected complicated HSP and more than one affected patient were further evaluated with the step 2 panel. This extended step 2 panel further delineate the genetic spectrum of rare autosomal recessive complicated HSP.


This study is focused on investigating the functional aspects of a recently identified gene, PTCHD1, and its disruption leads to Autism Spectrum Disorder and/or Intellectual Disability. Firstly, we sought to identify PTCHD1 splice variants contributing different coding sequences that might have higher or more specific expression in brain. We identified a new transcript skipping exon 2. This new transcript is predicted to encode a 542 amino acid protein in comparison to the 888 amino acid protein encoded by the PTCHD1 long isoform. It has just 4 transmembrane domains, and encodes a 62 kDa protein as compared to 12-transmembrane domains in the long isoform and a 101 kDa protein. We also found the presence of an additional exon upstream of exon 1. The quantitative expression analysis demonstrates that expression of the two new PTCHD1 transcripts is highest in human cerebellum as compared to the brain sub-regions and other tissues. The biological relevance of the new isoforms remains to be investigated. Initial expression data, indicates that these additional PTCHD1 transcripts are expressed chiefly in the brain, thus, these transcripts and encoded isoforms may be more relevant to autism and ID. PTCHD1 is structurally similar to the Hedgehog (Hh) signaling pathway receptors PTCH1 and PTCH2. The Hh pathway plays an important role in embryonic development and adult stem cell functioning. To establish the involvement of PTCHD1 in Hh pathway, expression analysis was done with Hh pathway genes and putative PTCH1 partners. The quantitative expression analysis with over expressions PTCH2, GDF9C, Dlx5, Lef1, Gli1 and Gli2 in mouse brain, revealed that expression of PTCHD1 is restricted to the Hh pathway. Expression analysis was done with Hh pathway genes and putative PTCH1 partners. The qualitative expression analysis with over expressions of PTCH2, GDF9C, Dlx5, Lef1, Gli1 and Gli2 in mouse brain, revealed that expression of PTCHD1 is restricted to the Hh pathway.

1234F GABRB2 in Schizophrenia. H. Xue. Division of Life Science, Hong Kong University of Science and Technology, Hong Kong, China.

Deciphering the molecular basis of schizophrenia is essential to effective management of this devastating mental disorder. Over the past decade, my research group has focused on the basic research on schizophrenia etiology through the discovery and characterization of a schizophrenia-associated gene - GABRB2, coding for GABAA receptor β2 subunit. The association between schizophrenia and single nucleotide polymorphisms (SNPs) in introns 9 and 10 of GABRB2, first reported by my group, has been cross-validated for multiple ethnic groups. Functional impacts of the schizophrenia associated non-coding SNPs in GABRB2 have been demonstrated at both mRNA and protein levels, viz. genotype-dependent alterations in mRNA expression and splicing, and effects of genotypes on isoform ratios and electrophysiological attenuation of GABAA receptors. Through extensive molecular genetics, population genetics and evolutionary genetics characterizations, GABRB2 has been shown by us to be under strong positive selection, active recombination as well as genomic imprinting, likely contributing to a human lineage-specific insertion of an AluY transposable element. Their work on epigenetic regulation of GABRB2 revealed its developmental control and role in neuronal excitatory synapse formation. Most recently, they have also defined an extended GABRB2 association from psychotic disorders to social cognition. Our work has thus improved current understanding of schizophrenia at molecular level centered at GABRB2, which represents present one of the best characterized schizophrenia candidate genes.
Association of NTRK2 and emotional processing in healthy young subjects. K. Spalek1, A. Heck2, D. Coyne1, D.J.-F. de Quervain1, 4, A. Papassotiropoulos2, 3, 4. 1) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Basel, Switzerland; 2) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Basel, Switzerland; 3) University of Basel, Department Biozentrum, Life Sciences Training Facility, Basel, Switzerland.

NTRK2, also known as 'tropomyosin-related kinase' receptor B (TrkB), is one of two receptor types binding neurotrophins with high affinity. Its activation plays a central role in cell survival, differentiation and synaptic plasticity as well as neurotransmitter release. Several studies provide support for a relationship of NTRK2 and psychiatric disorders, e.g. depression, schizophrenia, addiction, eating and anxiety disorders. Expression studies in patients with psychiatric disorders and animal models of psychiatric diseases report reductions of TrkB expression in the brain. Furthermore, injection of TrkB agonist in mice' hippocampus induced long-term activation of TrkB and antidepressant effects whereas the injection of TrkB antagonist decreased TrkB activity and increased anxiety- and depression-like behaviours. In general, emotional dysregulation is a common denominator of many psychiatric disorders, and it can be assumed that NTRK2 is involved in this process. Based on the evidence for an influence of NTRK2 on emotional processing in psychopathology and the lack of studies investigating the role of NTRK2 in healthy young subjects, we investigated genetic associations between NTRK2 single nucleotide polymorphisms (SNPs) and emotional processing, measured by arousal ratings of emotional (negative and positive) and neutral pictures from the International Affective Picture System (IAPS) in a sample of 1'161 healthy young subjects. Our results show a significant association between two NTRK2 SNPs and emotional processing of positive pictures. Specifically, major allele carriers rated positive pictures as more arousing compared to non-carriers. We were able to replicate this association result in an independent sample of 822 healthy young subjects who performed the identical picture rating task. Our findings suggest a role of NTRK2 in emotional processing in healthy young subjects and might add useful information for the understanding of its role in psychopathology.

PRICKLE1 interaction with SYNAPSIN I reveals a role in Autism Spectrum Disorders. L. Paemka1, V.B. Mahajan1, J.M. Skeie1, L.P. Sowers3, S.N. Ehaideb1, P. Gonzalez-Alegra1, T. Sassaoka2, H. Tao2, A. Miyagi2, N. Ueno2, S. Wu1, B.W. Darbro2, P.J. Ferguson1, A.A. Pieper1, J.K. Britt1, J.A. Wemmie1, D.S. Rudd1, T. Wassink1, H. El-Shanti6, H.C. Melford6, G.L. Carvill1, J.R. Manak1, A.G. Bassuk1. 1) University of Iowa, Iowa City, IA; 2) Kitasato University School of Medicine, Japan; 3) Hospital for Sick Kids, Toronto, Canada; 4) National Institute for Basic Biology, Japan; 5) Shaffalah Medical Genetics Center, Doha, Qatar; 6) University of Washington, Seattle, Washington.

The frequent comorbidity of Autism Spectrum Disorders (ASDs) with epilepsy suggests a shared underlying genetic susceptibility; and several genes, when mutated, can contribute to both disorders. PRICKLE1 missense mutations were found to segregate with ASD, however the mechanism by which mutations in this gene might contribute to ASD is unknown. To elucidate the role of PRICKLE1 in ASDs, we carried out studies in Prickle1+/− mice and Drosophila, yeast, and neuronal cell lines. We show that mice with PRICKLE1 mutations exhibit ASD-like behaviors. To find proteins that interact with PRICKLE1 in the central nervous system, we performed a yeast two-hybrid screening with a human brain cDNA library and isolated a peptide with homology to SYNAPSIN 1 (SYN1), a protein involved in synaptogenesis, synaptic vesicle formation, and regulation of neurotransmitter release. Endogenous Prickle1 and Syn1 co-localize in neurons and physically interact via the SYN1 region mutated in ASD and epilepsy. Finally, a mutation in PRICKLE1 disrupts its ability to increase the size of dense-core vesicles in PC12 cells. Taken together, these findings suggest PRICKLE1 mutations contribute to ASD by disrupting the interaction with SYN1 and regulation of synaptic vesicles.

CMYA5, a candidate gene for schizophrenia: Expression in the brain and the effect of a functional variant on binding. A. Hsiung1, R. Shiang, X. Chen. Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA.

Schizophrenia is a devastating psychiatric disorder with a prevalence of approximately 1% and is characterized by delusions, hallucinations and deficits of cognition and emotion. Although many DNA variants are identified to associate with schizophrenia, the functional consequences of such variants that result in increased risk are unknown. In collaboration with a large group of investigators, we have found strong evidence that the CMYA5 gene is associated with schizophrenia in a two-stage study using more than 33,000 subjects (Chen et al., 2010). The non-synonymous single nucleotide polymorphism (SNP), rs10043986, is a functional variant that changes the highly conserved proline residue 4063 to leucine in the CMYA5 protein, myospryn. Current studies of CMYA5 focus on its role in skeletal and cardiac muscle, and no studies of CMYA5 in the brain and neuronal cells have been reported, the tissue most likely to be affected in patients with schizophrenia. We hypothesize that myospryn is expressed in the brain and this SNP affects the binding properties of myospryn to its binding partner, desmin intermediate filament. Using reverse transcription PCR, myospryn and desmin transcripts are shown to express in mouse brain regions. Western blotting analysis also confirms the protein expression of myospryn and desmin in the same brain regions. Next, we investigated whether the two variants of CMYA5 would change myospryn binding to desmin using a yeast two-hybrid assay (Y2H) and validating with surface plasmon resonance (SPR). The results show that both alleles of myospryn bind to desmin. A yeast two-hybrid assay shows that the Pro allele has significantly weaker binding compared to the Leu allele, providing evidence that rs10043986 results in change of myospryn’s binding to desmin. The SPR result shows that the Pro allele has higher equilibrium dissociation constant than the Leu allele in the steady-state analysis, confirming the differential binding. In conclusion, expression of CMYA5 in the brain provides evidence that it has a functional role in the brain and is important for the schizophrenia pathophysiology, and rs10043986 affects the binding properties of myospryn to desmin and is a potential causal variant.

Involvement of non-Hsa21 genes and MicroRNAs provide etiological basis for abnormal phenotypes in pathogenesis of Down syndrome. A. Pathak, S.R. Phadke. Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India.

Down syndrome (DS), the most frequent genetic disorder leading to mental retardation is caused by partial or complete triplication of human chromosome 21 (Hsa21). The differential expression of genes on extrachromosomal Hsa21 in DS leads to characteristic features of the disease. In this study, we determined the differential gene expression and understand the molecular mechanisms underlying pathogenesis of DS. To analyze the differential gene expression and understand the molecular mechanism underlying pathogenesis of DS, we performed global gene expression profiling in blood samples of 14 DS and 4 normal subjects using human whole transcriptome microarray. The microarray analysis revealed total of 624 genes (195 upregulated and 429 down regulated) were differentially expressed in DS patients as compared to control. Out of the genes present on chromosome-21, a total of 210 genes were differentially expressed ranging from 1.5 to 5 fold compared to normal individuals. Genes expressed in pathological pathways such as apoptosis regulation, cell cycle regulation, signal transduction, cell maturation, and immunity showed dysregulation. Several genes localized on chromosome-21 such as APP, SOD1, DYPK1A, COL6A1 showed differential expression and these levels were conserved across all DS subjects. Interestingly, several non chromosome-21 genes such as RCAN3 (chromosome 1), ANK3 (chromosome 10), CDK17 (chromosome 12) etc., having roles in cardiogenesis, signal transduction and differentiation of neurons showed conserved levels of expression across across all DS subjects. Further, the role of microRNAs in the regulation of gene expression, global miRNA profiling was performed in 4 Down syndrome patients and 1 control using Affymetrix miRNA 3.0 array. Several Hsa21 miRNAs like miR-99a, let-7c, miR-125b-2, miR155, miR-802 showed overexpression effecting the regulation of genes involved in DS pathogenesis. The gene dosage hypothesis on chromosome-21 may partially explain the neurological and other symptoms but our results substantiate the involvement of genes localized across different chromosomes in pathogenesis of DS. Further, identification of microRNAs involved in DS pathogenesis may lead to identification of new therapeutic targets for DS. Our data lead to more systematic and improved understanding of molecular mechanism underlying the pathogenesis of the disease.
More than ER stress: molecular mechanism for misfolded PLP1 that impacts subcellular dynamics and clinical severity of Pelizaeus-Merzbacher disease. Y. Numata1, T. Morimura2, S. Nakamura3, E. Hirano1, S. Kure1, Y. Goto1, K. Inoue1. 1) Mental Retardation and Birth Defect Research, NCNP, Kodaira, Tokyo, Japan; 2) Dept. Pediatrics, Tohoku Univ. School of Med, Sendai, Japan; 3) Unit for Neurobiology & Therapeutics, MNRC, Shiga Univ. of Medical Science, Otsu, Japan.

Involvement of endoplasmic reticulum (ER) stress and the subsequent unfolded protein response (UPR) has been implicated in pathogenesis of multiple human inherited diseases, including Charcot-Marie-Tooth disease (CMT) and Pelizaeus-Merzbacher disease (PMD). PMD-causing mutant PLP1 accumulates in the ER and induces ER stress. However, the link between the wide clinical severity of PMD and the cellular response induced by mutant PLP1 remains largely unknown. Here we identified that misfolded mutant proteins also impact global subcellular dynamics and cellular environment. We found that expression of mutant PLP1 in HeLa cells, M03.13 oligodendrocyte cells, and primary oligodendrocytes depletes major ER chaperones with a KDEL (Lys-Asp-Glu-Leu) motif from the ER. This can be detrimental to cells because accumulation of misfolded proteins in the ER generally up-regulate ER chaperones to alleviate ER stress. The PLP1 mutants also induced fragmentation of the Golgi apparatus (GA). These organelle changes are more prominent in cells expressing severer mutant, potentially associated with disease phenotype. Similar changes are also observed in CMT-causing MPZ mutant that triggers ER stress, suggesting a common molecular mechanism among these disease genes. Moreover, PLP1 mutants inhibited global trafficking of secretory and membrane proteins, possibly leading to deleterious cellular environment. Notably, we found that KDEL receptor, which is critical in retrotransport of ER chaperons, was also dislocated in the ER, possibly causing the depletion of ER chaperons. The fact that inhibition of GA to ER trafficking by brefeldin A induced similar cellular phenotypes supported this hypothesis. Altogether, we propose that misfolded disease-causing mutant proteins not only induce ER stress and trigger UPR, but also cause depletion of ER chaperones, GA fragmentation and protein trafficking deficit, further contributing to the pathogenesis of inherited ER stress-related diseases and their disease severity.

Potential regulatory functions of late-onset Alzheimer’s disease associated variants. S.L. Rosenthal1, M.M. Barmada X. Wang2, F.Y. Demircol1, O.L. Lopez1, M.I. Kamboh1,2,3. 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA; 3) Alzheimer’s Disease Research Center, University of Pittsburgh, Pittsburgh, PA, USA.

Late-onset Alzheimer’s disease (LOAD) is a multifactorial neurodegenerative disorder resulting in loss of cognitive and executive function and ultimately death. It currently affects over 5 million individuals in the United States alone. To date, ten loci (APOE, BIN1, CLU, EPHA1, PICALM, CD2AP2, CD33, ABCA7, MS4A4A/MS4A6E, and CR1) have been implicated as risk loci for LOAD. Despite a number of associated variants located in these loci, the identified associations show only modest effect sizes, with APOE as the sole producer of a considerable odds ratio. Given its complex nature, LOAD is a disease about which we stand to gain substantial insight by examining how these variants may alter transcription and subsequently, the potential regulatory function of previously published genome-pathology and phenotype. This study aims to effectively utilize the newly identified LOAD is a disease about which we stand to gain substantial insight by examining how these variants may alter transcription and subsequently, the potential regulatory function of previously published genome-pathology and phenotype. This study aims to effectively utilize the newly identified

A functional variation in the CHRNB3 promoter affects Parkinson’s disease risk and smoking. A. Bar-Shira1, M. Gana-Weisz3, Z. Gan-Or3, E. Giladi2, N. Giladi2, A. Orr-Urtreger1,3. 1) Genetic Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; 2) Movement Disorders Unit, Parkinson Center, Department of Neurology, Tel-Aviv Sourasky Medical Center, Tel-Aviv, Israel; 3) Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv, Israel.

Introduction: Parkinson’s disease (PD) is affected by various genetic alterations, environmental factors and aging. Interestingly, cigarette smoking is being considered as a protective factor, and it was suggested that nicotine and nicotinic-acetylcholine receptors (nAChRs) may play a role. Since the β3 nAChR subunit is depleted in the striatum of PD patients, we aimed to test the possibility that variations in the CHRNB3 gene, which encodes this subunit, are involved in PD. Methods: CHRNB3 was sequenced in 100 PD patients. The alteration that was found in the putative promoter of this gene was further analyzed in a cohort of 596 PD patients and 369 controls. Its effect on promoter binding and activity was studied in cellular models using Chromatin immunoprecipitation and Luciferase assays. Results: The minor G allele frequency was 0.31 and 0.26 among patients and controls, respectively (p=0.02), and carriers had an OR of 1.33 (95% CI=1.03-1.73) for PD. In addition, the minor allele was strongly associated with smoking in patients, as 48.4% of carriers reported smoking history compared to 32.6% of non-carriers (p<0.0001). The transcription factor Oct-1 binding was almost eliminated in lymphoblasts homozygous for the minor G allele, to only 6.5% percent of the binding in cells with the major A allele. Furthermore, the CHRNB3 promoter activity was reduced by 70%-96% in cells homozygous for the minor allele. Conclusions: The association between the CHRNB3 variation and smoking in PD patients and its functional role in the promoter may suggest a molecular link between PD, smoking and nAChRs. These findings may also raise the possibility that nicotine treatment in PD should involve genotyping and personal adjustment.

Identification of microRNA expression quantitative trait loci in the nucleus accumbens of human postmortem brains from alcohol dependent subjects and matched controls. M. Mandandi, V. Williamson, G. McMichael, B. Riley, K. Kendler, V. Vladimirou, Virginia Institute for Psychiatric and Behavioral Genetics, VCU, Richmond, VA.

Alcohol dependence (AD) is a chronic addiction disorder with heritable factors accounting for 60% of the risk. While genetic studies have identified numerous AD loci, the mechanisms underlying neuroadaptations to excessive alcohol consumption are unclear. MicroRNAs (miRNAs), a species of non-coding RNA, are abundantly expressed in the brain and predominantly function by down-regulating gene expression. Involved in normal brain development and function, miRNAs are also implicated in several neuropsychiatric disorders; however, only one study has assessed miRNA expression in human AD subjects and in the prefrontal cortex only. Here, we evaluated miRNA expression and function in the nucleus accumbens (NAC), a major dopaminergic brain region within the mesocorticolimbic pathway that is involved in drug-seeking motivation and reward. Expression quantitative trait loci (eQTL) are genomic loci that can regulate miRNA and/or mRNA expression; however, no such studies have been conducted in postmortem AD brain tissue. We hypothesized that AD will also impact the expression of those miRNAs and miRNAs differentially expressed in the NAC of AD subjects and controls. In our preliminary results we detected 240 differentially expressed miRNAs at p<0.05; after multiple testing correction we identified 85 discovery rate (FDR) <10%, and 29 miRNAs remained significant. Within this list were miRNAs reported to be involved in aging/neurodegeneration (hsa-miR-1538, -516, -487a) and neurodevelopment/function (hsa-miR-371, -154, -247). Interestingly, hsa-miR-154 is also reported to be associated with mesocorticolimbic pathway modulation in opiate addiction. To better understand the mechanisms contributing to the miRNA expression, we identified AD-associated eQTLs affecting miRNA expression. We detected 275 cis-eQTLs (within 1 megabase) for all 29 miRNAs (FDR <10%). Several of these cis-eQTLs are located in functionally relevant sites including potential target sites. To further explore the specific mechanisms driving AD, miRNA expression data derived from the same sample will be integrated with the genetic and miRNA data to 1) identify miRNA/mRNA AD-specific interactions and putative miRNA target sites. Further, we will test the possibility that variations in the UTR or other regulatory elements of these interactions. Subsequently, weighted gene co-expression network analyses (WGCNA) will be performed to identify specific gene network modules under the control of AD-implicated miRNAs.
1247W
Behavioral and neurochemical characterization of mutant mice lacking Lphn3, a gene implicated in ADHD and addiction. D. Wallis1, C.A. Orsini2, B. Sellow1. 1) Biochemistry and Biophysics, Texas A&M University, College Station, TX; 2) Department of Psychiatry McKnight Brain Institute University of Florida College of Medicine Gainesville, FL 32610.
The latrophilin 3 (Lphn3) gene has been linked to susceptibility to attention deficit hyperactivity disorder (ADHD) and vulnerability to development of addiction. This suggests that this gene may be a genetic biomarker of these disorders and may lead to more selective therapeutic targets for treating them. However, little is known about the function of this gene. To characterize the function of the Lphn3 gene, we generated mutant mice. We then performed neurochemical, pharmacological, and behavioral assays to assess Lphn3 function. Four to 6 week old male Lphn3 mutant mice had significantly higher dopamine and serotonin levels in the dorsal striatum than their wild type (WT) counterparts. Given that elevated striatal dopamine is associated with locomotor hyperactivity in other mouse mutant lines, we evaluated locomotor activity in an open field arena. As expected, Lphn3 mutant mice displayed significantly more horizontal activity than both WT and heterozygous mice. Additionally, when Lphn3 mutant mice were administered acute i.p. injections of cocaine (20 mg/kg), they showed a significant elevation in locomotor activity relative to WT mice. In a separate cohort of male and female mice, we explored the contribution of the Lphn3 gene to reward-seeking behavior by assessing instrumental responding (lever pressing) for food pellet rewards under various fixed ratio (FR) schedules of reinforcement. Mice were first trained on a FR1 schedule for five 30 min sessions, after which they were tested on FR3, FR10, FR20 and FR40 schedules (one schedule/session). Lphn3 mutant mice displayed significantly greater instrumental responding for food than WT mice, particularly under high response ratios. Finally, we evaluated performance on rotorod assays and found a statistically significant difference between WT and null Lphn3 females with mutants having decreased latency to fall. Further, the WT mice were more likely to do a passive rotation on the rod, and the mutants were more likely to fall off. Together, these findings are consistent with a role for Lphn3 in regulating behavior, and show that a loss of gene function results in increased reward-seeking, possibly via enhanced striatal monoamine signaling. Our results demonstrate that Lphn3, a gene involved in gene function and neuropathological abnormalities that develop during childhood onset neurodevelopmental and genetic disorder characterized by chronic motor and phonic tics. Recent studies have indicated that the signaling. Current work is focused on characterizing Lphn3 mutant mice in the AVA interneuron can block forgetting similar to deletion of the msi-1 gene. The role of msi-1 on forgetting is also reflected in add-1 mutants by the strong and persistent consolidation of GLR-1 containing synaptic size increase induced by associative learning. Finally, we demonstrate that GLR-1 signaling regulates both action capping through the activation of add-1 and inhibition of the Arp2/3 complex translated to neuronal development and regulation of the structure and complexity of the actin cytoskeleton in neurons.

1248T
Induced Pluripotent Stem Cells to Model Tourette’s Syndrome. L. Deng1,2, N. Sun1,2, Z. Pang2, RA. King4, M. Sheldon1,2,3, JC. Moore1,3, RP. Hart4,5, M. Konosaki1,2, M. State6, GA. Heiman1,2, JA. Tischfield2,3, 1) Dept Genetics, Rutgers Univ, Piscataway, NJ; 2) Human Genetics Institute of New Jersey, New Brunswick, NJ; 3) Yale Child Study Center, Yale University School of Medicine, New Haven, CT; 4) Department of Psychiatry , UCSF, San Francisco, CA. Much of our current knowledge about the central nervous system (CNS) and neural function in patients with neuropsychiatric disorders has been obtained from relatively degraded postmortem brain. The inability to sample live CNS tissues impedes our progress to understand possible alterations in gene function and neuropathological abnormalities that develop during the course of the disorder. The rapid growth of iPS technology has turned somatic cells into multipurpose basic and clinical research tools and opened new windows for modeling human diseases. Tourette syndrome (TS) is a childhood onset neurodevelopmental and genetic disorder characterized by chronic motor and phonic tics. Recent studies have indicated that the histrionicaminergic neurotransmission and POLR3A and POLR3B transcription efficiency, leading to a decrease in the expression of key transcripts for the development and maintenance of CNS white matter. We designed a custom microarray with probes for all known Pol III transcripts to compare expression levels between HML.

1250W
Transcriptional impact of POLR3A and POLR3B mutations in Pol III–regulated hypomyelinating leukodystrophies. K. Choquet1, S. Denuieu1, G. Boldina1, M-J. Dicaire1, D. Forget2, K. Guerrero2, B. Coulombe2, M. Teichmann2, G. Bernard3, B. Bras4. 1) Neurogenetics of Motion Laboratory, Montreal Neurological Institute, McGill University, Montreal, Canada; 2) INSERM U869, Institut Européen de Chimie et Biologie, Université Bordeaux Segalen, Bordeaux, France; 3) Institut de Recherches Cliniques de Montréal (IRCM), Montreal, Canada; 4) Pediatric Neurodegenerative Laboratory, Department of Pediatrics, Neurology and Neurosurgery, Division of Pediatric Neurology, Montreal Children's Hospital, McGill University Health Center, Montreal, Canada. Leukodystrophies are a heterogeneous group of neurodegenerative diseases characterized by abnormal white matter in the central nervous system (CNS). It has been established that mutations in POLR3A and POLR3B are an important cause of five clinically overlapping hypomyelinating leukodystrophies (HML). POLR3A and POLR3B encode the two catalytic subunits of RNA Polymerase III (Pol III), which synthesizes over 200 non-coding RNAs including transfer RNAs (tRNAs). We hypothesized that mutations in POLR3A and POLR3B could perturb expression of tRNAs leading to a decrease in the expression of key transcripts for the development and maintenance of CNS white matter. We designed a custom microarray with probes for all known Pol III transcripts to compare expression levels between fibroblasts of four POLR3A–confirmed cases and four controls. A relatively small group of transcripts were found to be downregulated (n=21), of which an even smaller number (n=5) were found to be more expressed in brain compared to other tissues. We further validated their downregulation by Northern Blots in a larger cohort comprising POLR3A- and POLR3B–confirmed cases as well as healthy controls. We also isolated Pol III from patients and controls fibroblasts’ nuclei and performed quantitative RT-PCR to confirm the microarray observations. Deregulated Pol III targets are further investigated to determine the role of these alterations in the pathophysiology of Pol III–related HML. Elucidating the molecular mechanisms responsible for this disease is essential to develop potential therapeutic approaches.
1251T
The NINDS Repository: A unique resource of patient-derived, highly characterized primary fibroblasts and induced pluripotent stem cells for neurodegenerative disease research. C.A. Pérez1,2, S. Heit2,3, S. Gandre-Babbe2,4, C. Rhoda2,4, K. Pancken2,4, M. Self2,4, K. Hodges3,4, M. Shutterland3, R.A. Corneveaux4, C. Tam1,4. 1) NINDS Repository; 2) Stem Cell Biobank; 3) Cell Culture Laboratory; 4) Coriell Institute for Medical Research, Camden, NJ. The National Institute of Neurological Disorders and Stroke (NINDS), Bethesda, MD.

Induced pluripotent stem cells (iPSCs) reprogrammed from patient-derived primary fibroblasts have become an increasingly utilized resource for the study of disease. The NINDS Repository has added over 200 iPSC lines from PD, ALS, HD, frontotemporal degeneration, or Alzheimer’s disease, as well as samples derived from neurologically normal controls. For certain affected individuals, the parental fibroblast, corresponding iPSC line, and whole blood DNA are available. It is important to note that all reprogrammed cells are from the same individual, and their specific characterization details describing publicly available iPSCs and fibroblasts are documented and clearly presented to potential requestors. The NINDS Repository serves as a unique and effective centralized resource for these iPSCs and fibroblasts and their critical phenotypic data, to basic and applied research investigators worldwide.

1251W
Investigation of maternal stress among women who carry the FMR1 premutation who are mothers of a child with fragile X syndrome: involvement of the corticotrophin-releasing hormone receptor 1 gene to impact maternal well-being. J.E. Hunter2, A. Loni1, D. Hamilton1, L. Shubeck1, A. Abramowitz2, J.F. Cubells3, S.L. Sherman1. 1) Dept Human Genetics, Emory Univ, Atlanta, GA; 2) Dept Clinical Psychology, Emory Univ, Atlanta, GA; 3) Dept Psychiatry and Behavioral Sciences, Emory Univ, Atlanta, GA.

The FMR1 contains a polymorphic CGG repeat which, once unstable, is capable of expansion across generations with maternal transmission. Women who carry a premutation allele of FMR1, defined as an allele with 55-199 repeats, are at risk of passing on a full mutation allele, defined as an allele with >200 methylated repeats, to their offspring. Methylated full mutation alleles result in the intellectual disability disorder, fragile X syndrome (FXS). We recently published a study among mothers of children with FXS who carry a premutation showing evidence that polymorphisms within the corticotrophin releasing hormone receptor 1 gene (CRHR1), a major regulator in the endogenous cortisol response to stress, moderated the impact of maternal stress on the severity of distress experience in social situations as measured by the Social Phobia and Anxiety Inventory (SPAI). The goal of the current study is to follow up on these results to determine the severity and sources of maternal stress associated with raising a child with FXS involved in this interaction. Preliminary results for 25 mothers indicate that the interaction between maternal stress and CRHR1 polymorphisms is associated with both child characteristics (Aberrant Behavior Checklist: Irritability subscale, p=0.04; Social Avoidance subscale, p=0.02) and the impact of having a child with an intellectual and developmental disability on the family (Questionnaire on Resources and Stress: Parent and Family Problems subscale, p=0.01). The results of this study will provide potential targets for maternal interventions to alleviate stress and maternal well-being.

1252F
Understanding the Risk Pathway from GABRA2 to Alcoholism: Effects of Personality, Brain, Development and Gene x Environment Interactions. M. Burmeister1,2, S. Villafuerte1,2, E.M. Trucco1,2, M. Heitzeg2, R.A. Zucker2. 1) Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI; 2) Department of Psychiatry, University of Michigan, Ann Arbor, MI.

SNPs within GABRA2 have long been known to be associated with alcoholism. To understand the path from a genetic variant to complex behavior, alcohol abuse, we used the Michigan Longitudinal Study (MLS), consisting of ~463 families, and recruited ~75 years ago. The MLS includes all of his siblings and parents, ~70% of fathers met criteria for alcoholism, plus neighborhood controls. Assessments were every 3 years, teenagers every year. Measures considered here include impulsiveness in adults (NEO-PI-R), MRI during a modified incentive monetary task (life-time alcohol problems, youth externalizing behavior (Young Self Report) and parental monitoring. GABRA2 genotypes were tested for association by regression analysis. Mediation was tested using bootstrapping procedures in AMOS. Growth mixture modeling (GMM) was used to separate trajectories of alcohol problems in twins. We identified three trajectory classes of externalizing behavior across adolescence, a low, a developmentally limited, and a high risk class. Parental monitoring but not genotype predicted lower levels of externalizing behavior. We observed a significant GABRA2 x parental monitoring interaction effect on trajectories membership. While A-carriers’ trajectory class membership was largely unaffected by parental monitoring, those with the risk (G) genotype were affected by parental monitoring, both positively and negatively. These and our previously published results demonstrate that subjects who can (A-carriers) and cannot (G-carriers) engage in adaptive behavior become alcohol addicted and have alcohol problems. Our study demonstrates how understanding of the path from SNP to final phenotypic outcome can be achieved in smaller, well characterized longitudinal samples, and how genetic findings can be pulled into the existing psychosocial literature, merging studies of nature with nurture to explain some aspects of human behavior.

1252T
Advanced paternal age is associated with earlier onset of schizophrenia in the affected siblings of multiplex families. W.J. Chen1,2,3,4, S.H. Wang1,2, C.K. Hsiao1,2, C.M. Liu1,2, H.G. Hwu1,2,5. 1) Institute of Epidemiology & Preventive Medicine, National Taiwan Univ, Taipei, Taiwan; 2) Genetic Epidemiology Core Laboratory, Center of Genomic Medicine, National Taiwan University; 3) Institute of Brain and Mind Sciences, College of medicine, National Taiwan University; 4) Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University.

Introduction: Advanced paternal age has been reported to increase the risk of schizophrenia. Early onset, an important component phenotype for schizophrenia, may disrupt the normal processes of maturation of the brain and cause neurodevelopmental deviance. This study aimed to explore if advanced paternal age is associated with not only increased risk of schizophrenia but also earlier onset of schizophrenia in multiplex families. Methods: A total of 1359 affected siblings with schizophrenia and 308 healthy siblings from 694 multiplex families recruited throughout Taiwan were included for this study. Probands and their first-degree relatives were interviewed using the Diagnostic Interview for Genetic Studies. Offspring’s paternal age was calculated as its father’s age minus the offspring’s age at the time of recruitment. We compared the paternal age of affected siblings with their counterparts of healthy sibling within each family using random-effect models to control for familial dependence. Then we compared the distribution of onset age of schizophrenia for eight paternal age groups (<20, 20–25, 25–30, 30–35, 35–40, 40–45, 45–50, and ≥50). Results: Increasing 1 year in paternal age of schizophrenia for eight paternal age groups (<20, 20–25, 25–30, 30–35, 35–40, 40–45, 45–50, and ≥50). Results: Increasing 1 year in paternal age was associated with an increased risk of schizophrenia, with an adjusted odds ratio of 1.12 (95% confidence interval 1.03-1.22). Using paternal age of 20–25 as the reference, advancing paternal age was associated with younger age at onset of schizophrenia. Advanced paternal age is associated with not only increased risk of schizophrenia but also earlier onset of schizophrenia in multiplex families. How the paternal age modifies the risk of schizophrenia in offspring with a genetic susceptibility remains unknown. Future research will elucidate the mechanisms through which advanced paternal age accelerates the onset of schizophrenia.

1252F
The goal of the current study is to follow up on these results to determine the severity and sources of maternal stress associated with raising a child with FXS involved in this interaction. Preliminary results for 25 mothers indicate that the interaction between maternal stress and CRHR1 polymorphisms is associated with both child characteristics (Aberrant Behavior Checklist: Irritability subscale, p=0.04; Social Avoidance subscale, p=0.02) and the impact of having a child with an intellectual and developmental disability on the family (Questionnaire on Resources and Stress: Parent and Family Problems subscale, p=0.01). The results of this study will provide potential targets for maternal interventions to alleviate stress and maternal well-being.

1252T
Advanced paternal age is associated with earlier onset of schizophrenia in the affected siblings of multiplex families. W.J. Chen1,2,3,4, S.H. Wang1,2, C.K. Hsiao1,2, C.M. Liu1,2, H.G. Hwu1,2,5. 1) Institute of Epidemiology & Preventive Medicine, National Taiwan Univ, Taipei, Taiwan; 2) Genetic Epidemiology Core Laboratory, Center of Genomic Medicine, National Taiwan University; 3) Institute of Brain and Mind Sciences, College of medicine, National Taiwan University; 4) Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University.

Introduction: Advanced paternal age has been reported to increase the risk of schizophrenia. Early onset, an important component phenotype for schizophrenia, may disrupt the normal processes of maturation of the brain and cause neurodevelopmental deviance. This study aimed to explore if advanced paternal age is associated with not only increased risk of schizophrenia but also earlier onset of schizophrenia in multiplex families. Methods: A total of 1359 affected siblings with schizophrenia and 308 healthy siblings from 694 multiplex families recruited throughout Taiwan were included for this study. Probands and their first-degree relatives were interviewed using the Diagnostic Interview for Genetic Studies. Offspring’s paternal age was calculated as its father’s age minus the offspring’s age at the time of recruitment. We compared the paternal age of affected siblings with their counterparts of healthy sibling within each family using random-effect models to control for familial dependence. Then we compared the distribution of onset age of schizophrenia for eight paternal age groups (<20, 20–25, 25–30, 30–35, 35–40, 40–45, 45–50, and ≥50). Results: Increasing 1 year in paternal age was associated with an increased risk of schizophrenia, with an adjusted odds ratio of 1.12 (95% confidence interval 1.03-1.22). Using paternal age of 20–25 as the reference, advancing paternal age was associated with younger age at onset of schizophrenia. On average, the onset age was lowered for 1.5 years for the paternal age of 25–30 and 5.5 years for the paternal age of ≥50, with a P value of 0.04 for the trend test. Conclusion: Advanced paternal age was associated with an increased risk of schizophrenia in the offspring as well as earlier onset of schizophrenia in affected siblings. These findings derived from multiplex families of schizophrenia imply that the influence of advanced paternal age may be due to that fathers bequeath more mutations as they age.
1255F
Parents and teachers report on different aspect of children's and adolescent's conduct disorder and hyperactivity/inattention behavior. X.W. Zhang1,2, P.C. Sham2, S.S. Cherry3, H.Q. Meng4, Y.X. Fu5, Y. Huang4, T. Li1. 1) Psychiatry, West China Hospital, Sichuan University, China; 2) Psychiatry, Hong Kong University, Hong Kong; 3) Genome Research Centre, Hong Kong University, Hong Kong; 4) Mental Health Center, First Affiliated Hospital, Chongqing Medical University, China. Background: Either conduct disorder or hyperactivity/inattention problem poses considerable burden on health care and education, and the co-occurrence bring more pressure and impairments for the children and their family. However, the etiology of overlapping is still unclear, and one of the reasons is assessment inconsistent in different cultural background. Method: Subjects were 433 twin pairs aged between 6 and 16 years from Prospective Twin Registry in Southwestern China, whose parent and teacher completed the Strengths and Difficulties Questionnaires. It both contained estimation on twins' conduct disorder and hyperactivity problem from different point of views. And then used the structure equation model to explore the relationship between conduct disorder or hyperactivity/inattention problem in children and adolescent. Results: Both in the bi-traits twin model and bi-raters twin model, biometric model was the best fitting one. According to this model, conduct disorder and hyperactivity behavior were correlated with phenotypic correlation 0.45-0.65, and genetics factors contributed majority (around 70%) to the covariance both for boys and girls on parent's view, whereas environmental also played important (around 55%) in teacher's view. Meanwhile, the results between parent's and teacher's information were not that consistence. The phenotypic correlation between two informants was less than 0.2. In bi-raters twin model fitting, rater bias model was rejected. In the best fitting model, the common environmental effect always contributed more on variance from teacher's point of view than from parent's point of view, so I assumed that they maybe observe the children in different occasions, or evaluate them without fixed reference group. Conclusion: conduct disorder and hyperactivity behavior are correlated with each other, substantially contributed by genetic factors, but not governed by the common specific underly-ing phenotype. The information from parents and teacher are disparity, and it suggests us to collect data from more than one informant when using questionnaires to evaluate conduct disorder or hyperactivity/inattention problem in China.

1256W
Association of APOE Polymorphism and Stressful Life Events with Dementia in the Pakistani Population. M. Chaudhry1, S. Hasnain1, B. Smitz2, X. Wang3, D. Winger3, L. Wang4, S. Rosenthal5, F.Y. Demiric3, M.I. Kamboh1. 1) Department of Microbiology and Molecular Genetics, University of the Punjab, Lahore, Pakistan; 2) Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA; 3) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA; 4) Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, PA, USA. Dementia is a major public health problem worldwide. Alzheimer's disease (AD) is a major form of dementia and the APOE*4 allele is an established genetic risk factor for AD. Similarly, stressful life events (SLE) are also associated with dementia. The objective of this study was to examine the association of APOE*4 and SLE with dementia in a Pakistani population, which to our knowledge has not been reported previously. We also tested for an interaction between SLE and APOE*4 and the risk for dementia in this sample. A total of 176 subjects (61 cases and 115 controls) were recruited for this study. All pre-diagnosed cases and healthy controls were then interviewed to assess cognition, co-morbidities, history of SLE and possible differences in demographics. Blood samples were also drawn and genotyping for the APOE polymorphism (E2/E3/E4) was performed. The APOE*4 and stressful life events were each independently significantly associated with the risk of dementia. The odds ratios (ORs) for APOE*4 carriers and SLE were 2.81 (95% CI: 1.26-6.21; P=0.011) and 1.008 (95% CI: 1.004-1.012; P=1.15E-05), respectively. The gender stratified analysis revealed that APOE*4 and SLE were independently associated with dementia in males but not females. However, we did not find a significant interaction between APOE*4 carrier status and stressful life events in affecting the risk of dementia (P=0.677). Although the sample size of this study was small, the established association of APOE*4 with dementia was confirmed the first time in the Pakistani population. Furthermore, SLE was also found to be a significant predictor for dementia in this population. Our study also emphasizes the need to improve mental health facilities for older people in Pakistan and to facilitate future dementia research.

1257T
Association between MAOA and aggressive behavior in adolescents receiving the pharmaceutical treatment lisdexamfetamine dimesylate for ADHD symptoms. K.A. Nelson, M.S.1, A. Banker2, P. Huizenga, H.T. (ASCP)1, lATL1, S. Weaver, R.N.1, T. Jung1, E.A. Ehli, R.N., M.S.1, T.J. Soundy, M.D.1, K. Bohlen, PharmD3, Y. Hu, Ph.D.1, S.S. Cherny, Ph.D.1, 2, 3. 1) Avera Institute for Human Genetics, Avera McKennan Hospital and University Health Center, Sioux Falls, SD; 2) University of South Dakota, Sanford School of Medicine, Sioux Falls, South Dakota; 3) Department of Psychiatry, University of South Dakota, Sanford School of Medicine, Sioux Falls, South Dakota. Attention deficit hyperactivity disorder (ADHD) is one of the most common childhood neuropsychiatric disorders, affecting 5.29% of children worldwide. ADHD symptoms manifest as severely disruptive behaviors of inattention, hyperactivity, and/or impulsiveness. The pharmaceutical lisdexamfetamine dimesylate (LDX [Vyvanse®]), a prodrug stimulant prescribed for ADHD symptoms, is thought to alleviate symptoms by inhibiting the dopamine and norepinephrine reuptake pathways. Monoamine oxidase A (MAOA) is an X-chromosome linked gene that catalyzes the degradation of dopamine and norepinephrine. Wildtype alleles consist of 3.5 or 4 repeats (high enzymatic activity alleles). The genotypes of 2, 3, or 5 repeats (low enzymatic activity alleles) have been associated with increased aggression in males. This study looked at adolescents being treated for ADHD symptoms with LDX. Nearly half of the study individuals discontinued the medication due to aggressive behavior. Here, we hypothesized that individuals with the low enzymatic activity MAOA alleles would be the individuals more likely to discontinue the LDX due to aggressive behavior than individuals with the higher MAOA enzymatic activity genotypes. The study sample included 73 adolescents averaged 12 (6–18) years in age. The final study sample was 85% male and 90% Caucasian. All individuals were being treated within a continuation clinic setting and were monitored by licensed child psychiatrists for ADHD symptoms. After the ADHD diagnosis, each individual at one time during their treatment, received LDX medication. Each child’s buccal cell DNA was used to perform the genotyping of the MAOA 30-bp polymorphism. We found that individuals with high MAOA enzymatic activity alleles were significantly more likely to discontinue LDX due to aggression than individuals with the low enzymatic activity MAOA alleles (OR=0.3083 [95% CI: 0.11, 0.86] p=0.02). Stimulant medications are the first-line pharmacological treat-ment for children with ADHD. Although these treatments are very effective for the majority of its users, there are patients who can experience considerable adverse side effects. This study has shown a significant association between the discontinuation of LDX due to aggressive behaviors in adolescents with the high MAOA enzymatic activity alleles (3.5 and 4).
1258F

Psychosocial factors are correlated with gender-sensitive differences in relative telomere length (RTL) in Han Chinese university undergraduates. O. Yin1, M. Monakhov2, X. Zhang1, P.S. Lau1, S.H. Chew1, R.P. Ebstein1. 1) Pediatrics, National University Singapore, Singapore; 2) Dept. of Economics, National University Singapore, Singapore; 3) Business School, National University of Singapore, Singapore; 4) Psychology Dept., National University of Singapore.

Telomeres (TTAGGG repeats) cap the ends of chromosomes and numerous studies suggest that telomere length (TL) is a reliable index of cellular and tissue aging. More recently social stress has been shown to predict decreases in TL but much remains unknown about the psychosocial factors, especially in early adulthood, that modulate TL. To provide a better understanding of factors that influence TL and by implication aging and health, leukocyte DNA from 991 NUS Chinese undergraduates (52.4% Female) were measured for Relative TL (RTL) and compared to house-keeping β-Hemoglobin gene using the method of Cawthon (2009). A significant difference (T-Test = 0.0001 ; 989 d.f) in RTL by gender was observed and women, as previously reported, show greater RTL than male students. For women only, increased risk proneness over moderate gain, measured by incentivized behavioral economic tasks, was associated with shorter RTL (coef. =-0.0102; P = 0.040). Delay discounting was also associated in women only with reduced RTL (coefficient =-0.0089; P = 0.010). Impatient women have shorter telomeres. We also observed a relationship in women between RTL and scores on the NEO-PI-3 Neuroticism-Anxiety facet (coef =0.0072, p=0.045). Higher anxiety scores predict greater RTL in women. Intriguingly, for males and not females, students from families with incomes below the unofficial poverty line were characterized by reduced RTL (T-test = P = 0.022). Again in males only, NEO-PI-3 Agreeableness scores are associated with greater RTL (coef=-0.0208, p=0.017). Agreeableness is a personality trait suggesting the individual is kind, sympathetic, cooperative, warm and considerate. The current study examined the role of risk attitude, personality and socio-economic status on aging at the cellular level indexed by RTL. These psychosocial and behavioral characteristics appear to affect RTL but differently in men and women. Such gender-sensitive responsiveness in RTL to psychological and social environmental cues, that are already present in early adulthood in a non-clinical population, are suggested to have important implications for differential aging in men and women over the course of the lifespan.

1260T

The impact of the metabotropic glutamate receptor and other gene family interaction networks on the autism spectrum disorders. D. Hadley, Z. Wu, C. Kao, A. Kini, A. Mohamed-Hadley, K. Thomas, L. Vazquez, H. Qiu, F. Menitch, R. Pellegrino, C. Kim, J. Glessner, H. Hakonarson, Autism Genome Project Consortium. The Children’s Hospital of Philadelphia, Center for Applied Genomics, 3615 Civic Center Boulevard, Philadelphia, PA 19104. As multiple defective duplicated genes have been discovered to contribute to the etiology of the autism spectrum disorders (ASDs), we prioritized pathways robustly enriched for structural defects of duplicated genes and their gene family interaction networks (GFINs) using a novel network permutation test. Across 6,742 patients with ASDs relative to 12,544 neurologically normal controls, the metabotropic glutamate receptor (GRM) GFIN was significantly enriched for structural defects (P <= 2.40E-09, 1.8-fold enrichment) in ASDs and previously observed to impact attention deficit hyperactivity disorder (ADHD) and schizophrenia. The MAX dimerization protein (MxD) GFIN was also significantly enriched (P <= 3.83E-23, 2.5-fold enrichment) and known to regulate voltage independent calcium-activated action potentials at the neuronal synapse. Taken together, our genome-wide focus on GFINs illustrates the respective contributions of many different biological pathways in the etiology of ASDs.

1261F

The potential use of uncommon variants in GWAS data sets of pedigrees with autism spectrum disorders. H.Z. Wang1, H. Qin1, K. Ahn2, W. Guo3, Y. Y. Shugart1. 1) Division of Intramural Research Program, National Institute of Mental Health, Bethesda, MD; 2) Childhood Psychiatry Branch, National Institute of Mental Health, Bethesda, MD. GWAS data sets are often overlooked in rare-variant analysis, for the simple reason that DNA chips only sample a small subset of the genome. However with collapsing methods, rare-variants are linearly combined and analyzed as a group rather than individually, and rare variant analysis can further be enhanced using pedigree based methods, since variants rare in the population are enriched within families. We therefore explored the possibility of analyzing rare variants in GWAS data sets using such methods. To test this hypothesis we used FBAT-Rare, a pedigree-based collapsing method, and collapsed rare variants in genes using a frequency based weighing scheme. We applied this method to the Autism Genome Project (AGP) data set consisting of 2665 trios with autism spectrum disorder (ASD) afflicted offspring. We focused our efforts on CNTNAP2, a gene that encodes a member of the neurexin family which functions in the vertebrate nervous system as cell adhesion molecules and receptors. It is positively identified by multiple research teams, including Anney et al. (2012), the authors of the AGP data itself, to be strongly associated with ASD as well as other neurological development disorders. Using FBAT-Rare, we are currently able to detect an association of p=6.37*10^-5 for the European-only, high IQ male population for ASD, and p=2.77*10^-3 for strict autism. We realize the presented results do not meet the standard GWAS-wide significance level. However, our results appear to support the previous findings on CNTNAP2 reported by other investigators. More research work will focus on using several other analytical tools to re-analyze the AGP data set obtained from dbGap.
Genome-wide scan identifies candidate loci for gene-gene interactions affecting relationships between Aβ42, p-tau, and disease status. M.T.W. Ebendörfer1, P.G Ridge2, K. Boesche3, M. Bailey4, C. Crucchaga5, A. Goate6, S. Bertelsen7, C.D. Corcoran8, T. Maximoff9, J.S.K. Kauwe1, Alzheimer’s Disease Genetic Consortium. 1) Department of Biology, Brigham Young University, Provo, UT; 2) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 3) Department of Psychiatry, Washington University School of Medicine, St. Louis, MO; 4) Department of Mathematics and Statistics, Utah State University, Logan, Utah; 5) Center for Epidemiologic Studies, Utah State University, Logan, Utah; 6) Human Genetics Center, University of Texas, Houston, Texas.

Alzheimer’s disease (AD) is complex and affects 35.6 million people worldwide. Aβ42 and p-tau levels in cerebrospinal fluid are key biomarkers for the presence of pathological features of AD, Aβ plaques and neurofibrillary tangles, respectively. Like any complex disease, however, there are individuals that exhibit brain pathology indicative of AD but never develop clinical symptoms. In fact, more than 25% of individuals exhibit such pathology but show no cognitive decline. Likewise, there is considerable variation in the rate of decline in AD patients. These remarkable observations and other poorly understood attributes of AD are likely the result of undiscovered complex gene-gene interactions. To discover candidate loci for gene-gene interactions driving AD we used genotype data from 2,726 subjects for whom CSF biomarker levels have been measured to test for relationships quantitative trait loci (rQTL) that modify the relationships between (1) Aβ42 and p-tau; and (2) Aβ42 and disease case-control status. These rQTL are likely involved in gene-gene interactions but may be overlooked in typical association studies because they often do not have marginal effects on the individual traits whose relationship they modify. Specifically, we hypothesize that genetic modifications that modify these well-characterized relationships exist and explain, at least in part, variation in disease progression and the numerous individuals exhibiting known AD brain pathology without exhibiting clinical symptoms. We hypothesize that genetic factors that modify these relationships are likely to alter disease risk or age of onset. Scanning for rQTL in genome-wide SNP data we identified 3 genome-wide significant SNPs that may modify the relationship between Aβ42 and p-tau, and 2 genome-wide significant SNPs that may modify the relationship between Aβ42 and case-control status. rs807714 (intergenic between NAPAP1 and SNRPN, p < 1.7e-09), rs1036819 (intron of ZFAT and ZFAT-AS1, p < 4.2e-08), and rs7844573 (intron of ZFAT and ZFAT-AS1, p < 1.7e-09), rs1036819 (intron of ZFAT and ZFAT-AS1, p < 1.7e-09), rs10520400 (intron of PSD3 and exon of RPL35P6 pseudogene, 9.7e-08) may modify the relationship between Aβ42 and p-tau, while rs57216348 and rs59251601 (both in intron of CPLX2, p < 2.881-08) may modify the relationship between Aβ42 and case-control status. We concluded that the reported SNPs affect the relationships between (1) Aβ42 and p-tau; and (2) Aβ42 and disease case-control status, and that they may play an important role in AD etiology.
Our results suggest that the PRDM10 controls from 2,265 families) using the HapMap 2 imputed data. A family-based LOAD Study 3,828 subjects (1,840 Alzheimer disease cases and 1,988 controls) attempted to replicate our top findings in a family-based cohort, the NIA-LOAD. SNPs were significantly (p<10^{-5}) from the discovery GWA analyses were further evaluated (best SNP: rs1880916, p = 3.64 × 10^{-6}). No SNP met the initial cutoff for genotyped SNPs in the PDE11A containing 10) for BNT (best SNP: rs6590429, p = 1.81 × 10^{-7}) and four PRDM10 tests (WRAT) and the Trails B test (TRB) (0.30 to 0.72, p < 10^{-4}) after Boston, MA, USA.

Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; 5) Department of Biostatistics, Boston University School of Medicine, Boston, MA, USA; 3) Department of Biostatistics, Boston University School of Medicine, Boston, MA, USA; 4) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 2) Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA; 1) Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA; 4) Mount Sinai School of Medicine, USA; 5) Harvard Medical School, USA; 2) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 3) Department of Pediatrics, The Ohio State University; Nationwide Children’s Hospital, Columbus, OH; 4) Department of Neurology, Thomas Jefferson University, Philadelphia, PA; 5) Department of Biomedical Sciences, Cooper Medical School of Rowan University, Camden, NJ; 7) Division of Human Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 8) Department of Neurology and Pediatrics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 2) Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA.

Neuropsychological (NP) tests are commonly used to assess cognitive function and impairment in clinical settings. Recently, Genome-Wide Association (GWA) studies for general cognitive ability have combined tests representing domain-specific cognitive functions and failed to uncover genome-wide significant associations. We performed heritability and GWAS for 44 NP tests in the Rawer cohort, the Framingham Heart Study (FHS). The three most heritable NP tests were the Boston naming test with no cues (BNT), the Wide range achievement test (WRAT) and the Trails B test (TRB) (0.30 to 0.72, p < 10^{-4}) after adjustment for age, sex and race. Two SNPs (rs9847303 and rs4977861) genotyped SNPs were conducted separately for each test after adjustment and normalization from the FHS sample. The most significant associations were discovered with two genotyped SNPs in the PRDM10 gene (PR domain containing 10) for BNT (best SNP: rs6590429, p = 1.81 × 10^{-7}) and four genotyped SNPs in the PDE11A gene (Phosphodiesterase 11A) for WRAT (best SNP: rs1880916, p = 3.64 × 10^{-6}). No SNP met the initial cutoff for the significant association level p<10^{-5} with TRB. The association of 17 SNPs (p<10^{-5}) from the four NP tests were further evaluated with five additional NP tests. The four PDE11A SNPs were significantly associated (p<10^{-3}) with two more tests (best SNP: rs4335982 with Wechsler Adult Intelligence Scale- Similarities p=5.0 × 10^{-7}), while the association of the PRDM10 SNPs was limited to the BNT test (other tests: p > 10^{-3}). We also performed a Pan-ethnic GWAS, with 44 NP tests in the Genova/Eugenia/Nebraska/Virginia/San Francisco cohorts. Among these results, we observed, and in one case, we observed increased translation of LIMK2, a paralog of the WS-region Lim domain kinase 1 (LIMK1) gene. A profile of protein changes at the synapse is in progress. These data suggest a role for inherited translational differences in the expression of mRNAs in interpreting how CNVs alter the abundance of their encoded protein(s) and their secondary effects on the abundance of proteins expressed from other regions of the genome.

Genome-wide association study of epilepsy in a multi-ethnic cohort.

1266T Genome-Wide Association Study for Domain-Specific Cognitive Function. H. Milos Rasouly1, J. Chung1, R. Au3, L.A. Farrer1,2,3,4, G. Jun3, G. Jun3, 1) Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA; 2) Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA; 3) Department of Biostatistics, Boston University School of Medicine, Boston, MA, USA; 4) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 5) Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA.

Scanning for rare copy-number variation in 11,850 Swedish schizophrenia cases and controls revealed a surprising number of rare copy number variants associated with epilepsy, autism, and schizophrenia. XYLT1 also lies in an amplicon region containing genes encoding glutamate receptors, including MAPK signaling and in actin cytoskeletal regulation. However, to help close the gap to brain circuitry and cellular function, protein consequences of transcriptional hints must be elucidated both in WS and in emerging data on Autism and Schizophrenia. To do this, we examined the level of protein synthesis in WS-derived LCLs by using ribose profiling, a quantitative measure of translation that utilizes deep sequencing of ribosome-protected fragments. We analyzed the WS region transcripts by RNA-Seq and confirmed the expected 2-fold decrease in mRNA levels caused by the deletion. A ribose profiling experiment reports a previously unreported similar decrease at the level of actively translating ribosomes on these mRNAs, indicating that there is no compensation for the gene dosage effect at the level of translation for these mRNAs. However, perturbations in the translation of a single gene can play a role. This is observed, and in one case, we observed increased translation of LIMK2, a paralog of the WS-region Lim domain kinase 1 (LIMK1) gene. A profile of protein changes at the synapse is in progress. These data suggest a role for inherited translational differences in the expression of mRNAs in interpreting how CNVs alter the abundance of their encoded protein(s) and their secondary effects on the abundance of proteins expressed from other regions of the genome.

1266W Genome-wide association study of epilepsy in a multi-ethnic cohort. J.P. Bradfield1, Z. Wei2, C. Kim1, R. Chiavacci1, F. Mentch1, W. Lo2, M.R. Sperling4, D.J. Dlugos5, T.N. Ferrari6, R.J. Buono5, H. Hakonarson1,2,3,4,5, 1) Center For Applied Genomics, The Children’s Hospital of Philadelphia, Philadelphia, PA; 2) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 3) Department of Pediatrics, The Ohio State University; Nationwide Children’s Hospital, Columbus, OH; 4) Department of Neurology, Thomas Jefferson University, Philadelphia, PA; 5) Department of Pediatrics, University of Utah, Salt Lake City, UT; 2) Center for Neurogenetics and Human Behavior, Brain Institute, Department of Pediatrics, Univ. of Utah, Salt Lake City, UT.

Williams syndrome (WS) is a copy-number variation (CNV) disease caused by a recurrent 1.5 million base pair hemizygous deletion of 28 genes. The WS phenotype includes a specific cognitive profile and unique personality characteristics. Central to interpreting the genetic and neural contributions phenotype induced by these genomic changes is to go beyond the transcriptional consequences of the CNV associations and to examine how they might be closer to cellular functions, i.e., to determine abundance of WS encoded protein(s) and their downstream effects on the abundance and networks of proteins expressed from other regions of the genome. Our prior work showed striking, genome-wide, meaningful perturbations of the transcriptome response to the deletion in WS patient-derived lymphoblastoid cell lines (LCLs) using exon microarrays and RNA-Seq. We have shown that the WS deletion initiates a transcript level cis-effect that perturbs both WS region mRNAs, and mRNAs forming a network of interacting proteins and their associated pathways, including MAPK signaling and in actin cytoskeletal regulation. However, to help close the gap to brain circuitry and cellular function, protein consequences of transcriptional hints must be elucidated both in WS and in emerging data on Autism and Schizophrenia. To do this, we examined the level of protein synthesis in WS-derived LCLs by using ribose profiling, a quantitative measure of translation that utilizes deep sequencing of ribosome-protected fragments. We analyzed the WS region transcripts by RNA-Seq and confirmed the expected 2-fold decrease in mRNA levels caused by the deletion. A ribose profiling experiment reports a previously unreported similar decrease at the level of actively translating ribosomes on these mRNAs, indicating that there is no compensation for the gene dosage effect at the level of translation for these mRNAs. However, perturbations in the translation of a single gene can play a role. This is observed, and in one case, we observed increased translation of LIMK2, a paralog of the WS-region Lim domain kinase 1 (LIMK1) gene. A profile of protein changes at the synapse is in progress. These data suggest a role for inherited translational differences in the expression of mRNAs in interpreting how CNVs alter the abundance of their encoded protein(s) and their secondary effects on the abundance of proteins expressed from other regions of the genome.
1269T

Analysis of onset age in Late-onset Alzheimer Disease genome-wide association data identifies novel onset age loci and confirms the predominance of APOE. We identified 17 loci associated with AAO on chromosome 1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, and 22 (rs965857, P = 3.30×10^{-9}). These included chromosome 18q21.33 (rs12956834, \(P = 1.62 \times 10^{-9}\)) with one set of signals nearing genome-wide statistical significance in a chromosome 13q33.3 biological candidate region. The present study also identified several other novel variants with suggestive evidence of association with age at onset of AD in large high-risk families from the Collaborative Study on the Genetics of Alcoholism and the Family Study of Late-onset Alcohol Dependence (FSLA) at 14q31 (rs1042772, \(P = 7.09 \times 10^{-8}\)).

1270F

Mapping Alzheimer Spectrum Disorder Susceptibility Loci in Hispanic or Latino Populations. Y. S. Park1,2, N. Dueker1, J. Jaworski1, I. Konidari1, P. L. Whitehead1, C. Bustamante2, J. L. Haines1,2, M. A. Pericak-Vance1,2, M. L. Cuccaro1,2, E. R. Martin1,2, 1) John P. Hussman Institute for Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA; 2) Dr. John T. Macdonald Dept. of Human Genetics and Genomics, University of Miami, Coral Gables, FL, USA; 3) Vanderbilt University School of Medicine, Nashville, TN, USA; 4) Vanderbilt University School of Medicine, Vanderbilt University, Nashville, TN, USA.

In individuals of Hispanic or Latino origin (Hispanics), currently the largest minority group in the US and projected to be nearly 30% of the US population by the year 2050, present a unique challenge in complex disease genetic studies due to their admixed genomic background. Autism spectrum disorder (ASD) is highly heritable disorder with strong evidence for genetic influence. Current epidemiologic evidence suggests that ASD may occur as often as 1 in 88 children. Most ASD studies have primarily focused on individuals of European ancestry, making it difficult to generalize results in other populations. Our study interrogated a total of 253,682 SNPs genotyped in 2,033 individuals in over 550 families from three independent cohorts: HIHG (University of Miami and Vanderbilt University), AGRE (Autism Genetic Resource Exchange) and SSC (Simons Simplex Collection). Families with either the proband or at least one first degree relative self-reported Hispanics were included in the analysis. To our knowledge, this is the largest GWAS targeted to study ASD in Hispanics. These individuals have not been studied collectively in any other GWAS. Genomic ancestry was investigated with EIGENSTRAT principal component analysis (PCA) using HapMap 3 and GATK Linked Allele Set (LAST) genotype data. Genotype data from South Florida Hispanic families were used to estimate additional loci with effects as great or greater than APOE contribute to variation in AAO. Preliminary analyses confirmed association of APOE regional variation with AAO (rs965857, \(P = 3.30 \times 10^{-9}\)). Variants at several other LOAD risk loci also demonstrated statistically significant associations with AAO (\(P < 0.005\), including rs6701713 in CR1 (\(P = 0.00717\), rs7561528 in BIN1 (\(P = 0.00478\), rs561655 in PICALM (\(P = 0.00223\)). Burden analyses demonstrated that APOE contributes to 3.1% of variation in AAO (\(R^2 = 0.220\)) whereas the other nine genes contribute to 1.1% of variation (\(R^2 = 0.200\)) over baseline (\(R^2 = 0.189\), excluding study-specific effects. Secondary analyses of genomewide association with AAO performed among 10 ADGC case-control datasets (excluding cohort and family studies due to ascertainment differences) identified several regions with multiple SNPs showing suggestive associations (\(P < 10^{-5}\)). These included chromosome 1q21.33 (rs1256834, \(P = 1.62 \times 10^{-9}\)) and 9p13.3 (rs1735661, \(P = 2.49 \times 10^{-7}\)) with one set of signals nearing genomewide statistical significance in a chromosome 13q33.3 biological candidate gene MYO16 (rs9521011, \(P = 7.62 \times 10^{-10}\)). We confirmed the association of APOE variants with AAO among LOAD cases, and observe associations with AAO in CR1, BIN1, and PICALM. In contrast to earlier hypothetical modeling, we show that the combined effects of other loci do not exceed the effect of APOE on AAO, and if additional genetic contributions to AAO exist, they are likely very small individually.

1271W

Genome-wide association study of age at onset of alcohol dependence in large COGA families. M. Kapoor1, J. C. Wang1, L. Wetterholt2, S. Berenfeld1, J. Budd1, L. Nhung1, A. Agrawal1, V. Hesselbrock1, J. Rice1, J. Nurnberger1, L. Bierut1, T. Foroud1, A. Goate1, 1) Psychiatry, Washington University in St Louis, St Louis, MO, USA; 2) Indiana University School of Medicine, Indianapolis, USA; 3) University of Connecticut Health Center, Farmington, USA.

The age at onset of alcohol dependence (AD) is a critical moderator of genetic associations for alcohol dependence. The present study evaluated whether genetic variants can predict the occurrence of AD in large high-risk families from the Collaborative Study on the Genetics of Alcoholism (COGA). The primary analysis was performed in 2312 subjects from 118 of the largest and most densely affected families. Analyses were limited to the subjects who were regular drinkers. We tested for association between age at onset of AD and variants on the Illumina Omniexpress array using a Cox proportional hazards regression model after adjusting for the effects of birth cohort, gender and family structure. Non-dependent subjects were right censored at the age at last interview. A Schoenfeld residual test was also performed to test the proportional for the strongest signal. This family-based analysis identified a SNP, rs9847462 near micro RNA on chromosome 3, that is strongly associated (\(P = 1.03 \times 10^{-11}\)) with the age at onset of AD among regular drinkers. rs9847462 and other tagged SNPs were located under the region of linkage peak identified in our previous analysis. The results indicated that with each year of increase in age, carriers of the minor allele of rs9847462 were 1.5 times more likely to become AD than those homozygous for the wild-type genotype. By the age of 20 years, nearly 60% of subjects homozygous for the minor allele of rs9847462 variant were affected, compared to 20% of subjects homozygous for the major allele. This SNP was also associated with AD symptom count (\(P = 8.53 \times 10^{-7}\)) and DSM-IV AD (\(P = 3.04 \times 10^{-8}\)) in COGA families, but time to onset analysis significantly improved the power to identify these genetic variants. The present study also identified several other variants with suggestive evidence of association with age at onset of AD. These results indicate that there is a moderating effect of age on the risk of AD and time to onset analysis can increase the power to identify novel signals for AD.
1272T
Genetic determinants of the natural history of Alzheimer’s, X. Wang1, O.L. Lopez2,3, R.A. Sweet2,4, M.M. Barmada3, F.Y. Demirco1, M.I. Kamboh1,2,5. 1) HUMAN GENETICS, University of Pittsburgh, PITTSBURGH, PA; 2) Department of Neurology, University of Pittsburgh, Pittsburgh, PA; 3) Alzheimer’s Disease Research Center, University of Pittsburgh, Pittsburgh, PA; 4) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA. Alzheimer’s disease (AD) is a devastating neurodegenerative disease and is characterized by a gradual cognitive and functional decline. Alzheimer’s disease is a significant public health problem worldwide and the cost associated with AD is staggering. In order to effectively plan for the future care and treatments of AD patients, it is important to understand the factors that influence the natural history of AD. Although multiple factors have been identified that may affect the natural history of AD, including time to admission to nursing home (NH) and time to death, no studies have been conducted on the genetic determinants of the natural history of AD. In the present study, we have performed a genome-wide association study (GWAS) on 1,187 European American AD patients in order to identify genetic factors that may affect the time to NH admission and time to death. All patients were examined at the University of Pittsburgh Alzheimer’s Disease Research Center who had at least one follow-up evaluation (mean follow-up time 4.3 ± 2.7 years, range 0.8–18 years). Genotyping was performed using the Illumina Human Omni 1-Quad BeadChip and the analysis was conducted using Cox proportional hazards regression under an additive genetic model with adjustment for age, sex, education, baseline MMSE score, medication, psychosis and the first four principal components. The established genetic risk factor, APOE4, was not strongly associated with NH admission (P= 0.05) and death (P=0.15). However, we found multiple top significant signals for time to death located on chromosome 12, genes: ZUFSP (P=1.08E−07), RSPH4A (P=1.5E−07), and KPNA5 (P=6.45E−07). We also observed multiple suggestive loci for time to death of AD, including the top signal between the KDR and SDR5A3 genes on chromosome 18 (P=2.5E−07), ARAC1 on chromosome 10 (P=1.38E−06), SYNPO on chromosome 5 (P=2.36E−06), PPP1R3E on chromosome 14 (P=3.76E−06) and NEUROD4 on chromosome 12 (P=4.28E−06). Our data suggests that natural history of AD has genetic basis and this novel finding needs to be confirmed in independent studies.

1273F
A genome-wide association study of the response to cognitive behavioral therapy in children with anxiety disorders. J.R.I. Coleman1, K. Lester1, C. Curtis1, J.L. Hudson2, C. Creswell3, G. Brain4, T.C. Eley1. 1) MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London, UK; 2) Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia; 3) Winnicott Research Unit, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK.

Anxiety disorders are the most common psychiatric disorders, with a prevalence in adults of approximately 30%. Psychosocial treatments, including cognitive behavioral therapy (CBT), are the primary treatment modality for anxiety disorders in the United Kingdom. Response (at 6-12 month follow-up) for CBT is estimated at 65%, demonstrating heterogeneity. Evidence suggests individual differences in treatment response have a genetic basis. Preliminary evidence exists for beneficial effects of CBT on HttTLP3 SS genotype, and for the T allele of rs8330 (NGF gene), on CBT response at 3-12 month follow-up in a cohort of children with a variety of anxiety disorders (Eley et al., 2012, Lester et al., 2012). Such therapy-genetic candidate gene studies require robust genetic data.

Genetic samples were gathered from 1272 children (aged 7–13) diagnosed with anxiety disorders, undergoing CBT. The sample originated from ten different sites in the UK, USA, Australia and Western Europe. Clinical Severity Ratings were established at baseline, on completion of treatment, and at follow-up. DNA was extracted using a standard protocol, concentrated by filtration, and genotyped using the recently released Illumina HumanCore-Exome array, which assays roughly 250000 SNPs and 250000 exonic variants. The clinical outcome variable was primary disorder presence at follow-up. The SNPs most strongly associated with this outcome were presented. Additional measures include genome-wide complex trait analysis (GCTA) and quantitative trait heritability analyses. Further analyses will explore the association between SNPs and changes in Clinical Severity Ratings, and investigate response immediately post-treatment. This is, to our knowledge, the first GWAS of response to psychosocial treatment, and the first treatment response GWAS to be performed in anxiety disorder. Past analyses in related phenotypes such as antidepressant response suggest genetic determinants may be small. As the sample size of this study is relatively small, it is probably underpowered to detect such variants. However, GCTA could yield valuable insight; one study that found no SNPs at genome-wide significant levels for antidepressant response gave a SNP-heritability estimate of 42% from GCTA.

1274W
Genome-Wide Association Study of Shared Components of Reading Disability and Language Impairment. J.D. Eicher1, N.R. Powers1, L.L. Miller1, S. M. Ring2, J.R. Grun1, J.D. Pediatric Imaging, Neurogenomics, and Genetics Study. 1) Genetics, Yale University, New Haven, CT; 2) School of Social and Community Medicine, University of Bristol, Bristol, UK; 3) Pediatrics and Investigative Medicine, Yale University School of Medicine, New Haven, CT.

Written and verbal language are inherently intertwined neurobehavioral traits vital to the development of communication skills. Unfortunately, disorders involving these traits—specifically reading disability (RD) and language impairment (LI)—are common and prevent affected individuals from developing adequate communications skills, leaving them at risk for adverse academic, social, economic, and psychiatric outcomes. Both RD and LI are complex traits that frequently co-occur, leading us to hypothesize that these disorders share genetic etiologies. To test this, we performed a genome-wide association study on individuals affected with both RD and LI in the Avon Longitudinal Study of Parents and Children. The strongest associations were seen with markers in ZNF385D (OR=1.81, minimum p=5.45 × 10−7). These associations were strengthened when examining LI cases (OR=1.62, minimum p=5.96 × 10−8). We replicated association of ZNF385D using a case-control study, confirming the finding using negative control data. Further analysis of the ZNF385D locus identified that genes ZUFSP (P=1.08E−07), RSPH4A (P=1.5E−07), and KPNA5 (P=6.45E−07) are associated with LI.

1275T
A genome wide association between migraine in bipolar disorder and Neurobeachin. K.K. Jacobsen1, 2, 3, S. Johansson1, 2, 3, C.M. Nievergelt4, 5, 6, T. T. Yang5, 6, 5, 4, 5, 6, 6, T.A. Greenwood1, 2, 3, H. B. Akiskal4, 5, 6, 7, B. E. Haavik1, 2, 3, 6, O. B. Fasmer1, 2, 3, 6, 7, J. R. Kelsoe4, 5, K. J. Oedegaard4, 5, Bigs Consortium. 1) Department of Biomedicine, University of Bergen, Bergen, Norway; 2) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 3) Department of Psychiatry, University of California San Diego, San Diego, USA; 5) Department of Psychiatry, VA hospital, San Diego, USA; 6) Division of Psychiatry, Haukeland University Hospital, Bergen, Norway; 7) Section for Psychiatry, Department of Clinical Medicine, University of Bergen, Bergen, Norway.

Background: Migraine is a common headache disorder, with a prevalence of approximately 12%. It is characterized by unilateral attacks, and in some cases accompanied by visual aura symptoms (1). Bipolar disorder is a mood disorder ranging from severe depression to mania, with migraine as a common comorbidity seen in 25–45% (2, 3). Bipolar disorder patients with migraine are more likely to have a worse outcome than those without migraine (4, 5). Methods: We performed a genome wide association analysis on 460 bipolar patients with self-reported migraine and 914 patients without migraine. The individuals are from the TGEN sample, a part of the Bipolar Genetics Study (Bigs). Replication was attempted in the GAIN sample, a separate part of Bigs. Results: We found a genome wide significant association between migraine in bipolar disorder and rs1160720, an intronic single nucleotide polymorphism (SNP) in NBEA (P-value 2.97×10−8, OR=1.50, 95% CI: 1.27-1.75). The SNP resides in a linkage disequilibrium block with several other associated SNPs, spanning several exons. We were not able to replicate our finding in the GAIN sample. Discussion: NBEA encodes the protein neurobeachin, which is involved in the transport of neurotransmitter receptors, among them glutamatergic receptors (5). This receptor system is implicated in both bipolar disorder and migraine (6). Our current findings implicate new clues to the cause of migraine in bipolar disorder, though further studies are needed to verify this association.
1276F

Genome-wide association study of HLA-DQB1*06:02 negative essential hypersomnia. SS. Khor1, T. Miyagawa1, H. Toyoda1, M. Yamashita1, Y. Kawamura1, H. Tani3, Y. Okazaki4, T. Sasaki5, L. Lin6, J. Faraco5, T. Rico5, Y. Honda1, M. Honda1,6, E. Mignot6, K. Tokunaga1. 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan; 2) Yokohama Clinic, Warakukai Medical Corporation, Yokohama 221-0835 Japan; 3) Department of Psychiatry, Mie University School of Medicine, Mie 514-8507 Japan; 4) Metropolitan Matsuwara Hospital, Tokyo 156-0057, Japan; 5) Graduate School of Education, The University of Tokyo, Tokyo 113-0033 Japan; 6) Stanford Center for Sleep Sciences and Medicine, Stanford University School of Medicine, 1050A Arastradero Rd., Palo Alto, CA 94304, USA; 7) Department of Somnology, Tokyo Medical University, Tokyo 160-0023, Japan; 8) Sleep Research Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo 156-8506, Japan.

Essential hypersomnia (EHS), a sleep disorder characterized by excessive daytime sleepiness, can be divided into two broad classes based on the presence or absence of the HLA-DQB1*06:02 allele. HLA-DQB1*06:02-negative EHS and narcolepsy with cataplexy in both Japanese and Caucasian populations. This is the first GWAS of HLA-DQB1*06:02-negative Caucasian hypersomnia patients and 1761 HLA-DQB1*06:02-negative Japanese healthy controls. A comparative study was also performed on 268 HLA-DQB1*06:02-negative Caucasian patients lacking the HLA-DQB1*06:02 allele and 562 Japanese healthy controls. Interestingly, rs10988217 showed a similar tendency in its association with essential hypersomnia.

1277W

Genome-wide Association Study and Admixture Mapping of Age-related Cognitive Decline in African Americans. T. Raj1,2,3, L.B. Chibnik1,2,3, C. McCabe1, B. Stranger4, H. Hendrië5, L. Barnes6, T. Forough6, J.M. Murrell1, D.A. Bennett6, K.S. Hall3, D.A. Evans9, P.L. De Jager1,2,3, 1) Department of Neurology, Brigham and Women's Hospital, Boston MA; 2) The Broad Institute, Cambridge MA; 3) Harvard Medical School, Boston MA; 4) Section of Genetic Medicine, University of Chicago, Chicago IL; 5) Department of Medicine, Indiana University, Indianapolis IN; 6) Rush Alzheimer Disease Center, Rush University Medical Center, Chicago, IL; 7) Department of Medical and Molecular Genetics, Indiana University, Indianapolis IN; 8) Department of Pathology and Laboratory Medicine and Indiana Alzheimer Disease Center, Indianapolis IN; 9) Department of Psychiatry, School of Medicine, Indiana University, Indianapolis, IN; 10) Department of Internal Medicine, Rush Institute for Healthy Aging, Rush University Medical Center, Chicago, IL.

Objective: To leverage genome-wide data from several cohorts so as to identify genomic determinants of age-related cognitive decline among older African Americans.

Data and Methods: We examined genome-wide genotyping data for 3,964 unrelated older African American (AA) subjects who had at least two repeated measures of cognition from five (CHAP, IIDP, ROS, MAP, and MARS) prospective community-based studies. We compared them to 2,703 unrelated non-Hispanic European Americans (EAs) from the ROS, MAP and CHAP studies. Within each cohort, individual cognitive tests were combined to form an aggregate measure of global cognition. We used linear mixed effects models to characterize individual paths of change in cognition, controlling for age, sex and education as fixed effects. After quality control, imputation was performed using the 1000 Genomes Project Phase I combined reference panels of EA and AA ancestry. Common variants were meta-analyzed across cohorts using weighted fixed-effect models. HAPMIX and MIXSCORE programs were used to estimate ancestry at each locus, and to test for association of cognitive decline with locus-specific proportion of ancestry.

Results and Conclusion: APOE e4 haplotype was strongly associated with rate of cognitive decline in AAs (P= 1.92 x 10−11) but the magnitude of effect was significantly weaker in EAs (P= 0.01) compared to EAs (P= 0.01). In AAs we replicated previously known loci in EAs including ABCA7 (P= 4.13 x 10−4), PICALM (7.35 x 10−4), and EPHA1 (P= 8.06 x 10−4). We discovered one genome-wide significant association at the TRPS1 locus (chr 9p21), and replicated that association in EAs (P(DIS) = 2.52 x 10−7; P(REP) = 0.97, P(JOINT) = 0.37 x 10−6). The top variant in the region is significantly associated with cis gene expression of TRPS1 in primary monocytes. We also found another variant at the TEK locus (chr 9q21) with suggestive evidence of association (P(DIS) = 4.10 x 10−4). Using genome-wide admixture mapping, we did not detect any significant admixture peak, suggesting that genetic factors alone may not explain the differences in age-related cognitive decline between AAs and EAs.
1278T Genome-wide association study of growth rate and energy status in lymphoblastoid cell lines. A.R. Sanders1,2, W. Moy1, H.H.H. Gøring1, J. Freda1, D. He1, S.L. Fuentes1, J. Duan1,2, P.V. Gejman1,2, Molecular Genetics of Schizophrenia (MGS) Collaboration. 1) Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL; 2) Department of Psychiatry and Behavioral Sciences, University of Chicago, Chicago, IL.

The determinants of growth rate and energy status in LCLs, whether genetic, environmental, or both, remain largely unknown, though growth rate and energy status vary across LCLs and strongly influence cellular phenotypes. We performed a GWAS using 2,060 samples from the European ancestry (EA) portion of the MGS case-control collection with transformation site, age, EBV load, sex, caseness, and ancestry PCs as covariates. For both growth rate and energy status, we replicated their genes (or if intergenic, the genes closest on each side), resulting in 429 genes (growth rate) and 383 genes (energy status). The genomic inflation factors for the growth rate (1.000) and the energy status (1.014) GWAS indicated little to no cryptic population substructure. We found some associations for growth rate with age (p=0.018), sex (p=1.7×10^-4), and EBV load (p=8.3×10^-5), and for energy status with sex (p=1.8×10^-4), caseness (p=0.0076), transformation site (p=0.024), and EBV load (p=0.0090), suggesting the utility of including such covariates in the analysis. We found no genome-wide significant association. For growth rate, the strongest association with additional support was with rs7750067 (p=5.9×10^-7), with supportive findings (six SNPs with 5×10^-8<p<5×10^-7) in this intergenic region, which contains a predicted microRNA (miRNA), ENSG00000268073. The nearest genes are SLC25A15 (P1), EYS and BAI3. EYS contains multiple epidermal growth factor (EGF)-like and LamG domains and is implicated in retinitis pigmentosa. BAI3 encodes a brain-specific angiogenesis inhibitor, which is an adhesion-G protein-coupled receptor and has been shown to control dendritic arborization growth and branching in cultured neurons. Both EYS and BAI3 are expressed in various tissues, including brain and blood. Pathway analyses revealed the most enriched GO-term for growth rate to be neuron projection (FDR=0.08), and for energy status two were significant: regulation of cell proliferation (FDR=0.007) and neuron differentiation (FDR=0.042). We describe the effect of epidemiological variables and of specific genomic loci on growth rate and energy status in LCLs. Our results have potential to lead to a better understanding of these important cellular traits. We will present data on the expanded sample of LCLs at the meeting, projecting to reach ~2.500 EA subjects, along with ~1,000 African American subjects.

1279F Distinct genetic contributions to Tourette Syndrome and Obsessive Compulsive Disorder as revealed by cross-disorder genome-wide studies. D. Yu1,2 on behalf of TS GWAS Consortium, TSAICG, and IOCDFGC. 1) CHGR, Massachusetts General Hospital, Boston, MA; 2) Stanley Center, Broad Institute, Boston, MA.

Investigation of the genetic relationship of phenotypically related heritable diseases is important for identifying both shared and distinct genetic risk factors and ultimately for providing molecular insights into diagnosis and treatment. Polygenic score analysis of genome-wide association study (GWAS) data is a powerful tool to examine aggregate cross-disorder genetic factors and ultimately for providing molecular insights into diagnosis and diseases is important for identifying both shared and distinct genetic risk loci. In addition, OCD in the presence of TS/CT may have different underlying genetic susceptibility compared to OCD alone.


Introduction: Attention deficit hyperactivity disorder (ADHD) is a highly heritable childhood onset neuropsychiatric condition that often persists into adulthood. Still, genetics of ADHD, and particularly ADHD in adults is largely unknown. The main purpose of this study was to perform a genome-wide scan of adult ADHD using the newly available Illumina HumanExome12v1 chip, and to evaluate the performance of this technology for genotyping common and rare variants. Materials and Methods: The analyses were carried out using DNA samples collected by the International Multicenter persistent ADHD Consortium (ImPACT). The evaluation of HumanExome12v1 chip was implemented on 2215 individuals, 25 of which were also whole exome sequenced using Roche-NimbleGen Sequence Capture EZ Exome v2 kit and paired-end 100nt sequencing on the Illumina HiSeq. All participants were genotyped on Illumina HumanExome12v1 chip. Genotypes were called in Illumina GenomeStudio V2011.1 software, with additional genotype assignments in zCall software. Performance of HumanExome12v1 chip as well as GenomeStudio V2011.1 software was assessed by direct comparison of data obtained from different DNA sources (blood and saliva), duplicate samples and sequenced genotypes. Mendelian consistency testing was also utilized. Sanger sequencing was used to validate rare variants with minor allele frequency <1%. Association testing was carried out as single marker logistic regression correcting for population substructure and gender in PLINK software. Results: We developed a quality control analysis pipeline for the variants of HumanExome12v1 chip using series of steps implemented in PLINK, GenomeStudio V2011.1 and zCall softwares. Overall, the performance of HumanExome12v1 chip was comparable to that of next generation sequencing. However, it is worth mentioning that mismatch rates were notably higher for rare variants with minor allele frequency <0.5% compared to the rest of the chip, therefore exact specifics will be discussed in details. Association testing in a total of 2600 adult ADHD cases and 7000 controls are currently being performed and will be presented. Conclusion: This study provides insight into the performance of the newly available HumanExome12v1 chip, comparing it to conventional quality control with special emphasis on rare variants. Being the largest systematic adult ADHD study to date, this study expected to shed new light on our understanding of adult ADHD genetics.
there is little evidence that AD risk genes have a strong influence.

Conclusions:

Genetics influence rate of decline, although...
268

Posters: Psychiatric Genetics, Neurogenetics and Neurodegeneration

1283W

A Noise-Reduction GWAS in Multiple Sclerosis. J.P. Hussman 1,2, A.H.
Beecham 1 , M. Schmidt 1 , J.L. McCauley 1, J.L. Haines 3, M.A. PericakVance 1, International Multiple Sclerosis Genetics Consortium. 1) John P.
Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Hussman Foundation, Ellicott City, MD,
USA; 3) Center for Human Genetics Research, Vanderbilt University Medical
Center, Nashville, TN, USA.
Genome-wide association studies (GWAS) have proven useful in identifying disease susceptibility genes in multiple sclerosis. However, GWAS have
focused on the testing of each single nucleotide polymorphism (SNP) independently. Given it is likely that the disease variants have not been directly
genotyped, tests that account for multiple flanking SNPs in linkage disequilibrium (LD) with the disease variants may increase the power to detect association. We have used the GWAS noise reduction (GWAS-NR) approach to
identify LD blocks which are significantly associated with multiple sclerosis.
As GWAS-NR uses a linear filter to identify genomic regions demonstrating
correlation among association signals in multiple data sets, we have divided
the genotype data from a previously published GWAS in a total sample of
9722 cases and 17376 controls into seven population specific strata from
Australia, Central Europe, the Mediterranean, Finland, Scandinavia, United
Kingdom, and the United States. Within each stratum, we adjusted for the
first five principal components from Eigenstrat. Of the 87838 LD blocks in
the GWAS data, defined by the confidence intervals algorithm, we have
identified 1434 blocks with p-value < 5.0E−02. A total of 226 of the 1434
blocks show p-value < 5.0E-08 using a modified version of the Truncated
Product Method, 138 being within the Major Histocompatibility Complex
(MHC) on chromosome 6 (28-34 MB). Of the remaining 88, 24 contain SNPs
previously identified in GWAS at a genome-wide level and 47 lie within 2
MB of previously identified SNPs. The remaining 17 are in regions novel to
previous GWAS efforts. In particular, four of the 17 blocks are on chromosome 1 in the region of FCRL2, FCRL3, and FCRL4. Previous studies
have found that the Fc receptor-like (FCRL) molecule may contribute to the
autoimmune disease process. Other noteworthy findings include blocks
within JAK1 and COL11A2. Like TYK2, which has been shown to be associated with multiple sclerosis, JAK1 is a non-receptor tyrosine kinase. Three
blocks in COL11A2, located within the MHC at ~33 MB, show significance
(p-values < 5.0E-08). This is of interest as our recent analysis of sequence
data in multiplex families also pointed to this gene, with one of the identified
variants being located within one of the LD blocks. These data support the
use of this approach to identify additional genetic factors contributing to
multiple sclerosis risk.

1284T

Genome-Wide Association Study of Alcohol Consumption in the Kaiser
Permanente/UCSF Genetic Epidemiology Research on Adult Health
and Aging Cohort. E. Jorgenson 1, L. Shen 1, T. Hoffmann 2,3, M. Kvale 3,
Y. Banda 2,3, D. Ranatunga 1, N. Risch 2,3, C. Schaefer 1, C. Weisner 1,4, J.
Mertens 1. 1) Kaiser Permanente Division of Research, Oakland, CA; 2)
Department of Epidemiology and Biostatistics, University of California San
Francisco, San Francisco, CA; 3) Institute for Human Genetics, University
of California San Francisco, San Francisco, CA; 4) Department of Psychiatry,
University of California San Francisco, San Francisco, CA.
Alcohol consumption is a common trait with a complex etiology. There is
strong evidence that both genetic and environmental factors play a role
in determining an individual’s alcohol use, including patterns of alcohol
consumption and the risk of alcohol dependence. A number of genomewide association studies of alcohol dependence and alcohol consumption,
utilizing samples of several thousand subjects, have identified association
signals that require further confirmation in independent samples. Here, we
report the results of a genome-wide association study of alcohol consumption
in the Kaiser Permanente/UCSF Genetic Epidemiology Research on Adult
Health and Aging Cohort (n=110,266), the largest study to date. We identified
genome-wide significant associations between SNPs in the ethanol metabolism pathway and measures of alcohol consumption. Among non-Hispanic
white subjects who report drinking alcohol, we observed an association
between rs1229984, which encodes an arginine to histidine change in the
amino acid sequence of alcohol dehydrogenase 1B (ADH1B), the major
metabolizing enzyme of ethanol, and a reduction of 1.29 drinks per week
per copy (p=1.2*10 −26). We also observed a similar effect, a reduction of
1.47 drinks per week (p=6.9*10 −5), in Latino subjects, and a more modest
reduction of 0.38 drinks per week (p=0.05) in East Asian subjects. The
smaller effect of rs1229984 on the number of drinks consumed per week
in East Asian subjects is due to the presence in East Asian populations of
rs671, a polymorphism that encodes an amino acid substitution from glutamine to lysine in aldehyde dehydrogenase 2 (ALDH2), the major metabolizing enzyme of acetaldehyde, the intermediate product in the ethanol metabolism pathway. In East Asian subjects, rs77768175, which is in strong linkage
disequilibrium with rs671, was associated with a reduction of 0.80 drinks
per week (p=0.008). More strikingly, subjects who carry the rs77768175 G
allele are considerably less likely to report any alcohol consumption (OR =
0.32, p=3.6*10 −58).

1285F

Transethnic genome-wide meta-analysis of Alzheimer disease in Alzheimer's Disease Genetics Consortium. G. Jun 1,2,3, J. Chung 1, J. Kozubek 1, C. Reitz 6, B.N. Vardarajan 6, K.L. Lunetta 2, J.L. Haines 7, M.A. PericakVance 8, R. Kuwano 9, R. Mayeux 6, G.D. Schellenberg 10, L.A. Farrer 1,2,3,4,5,
Alzheimer's Disease Genetics Consortium. 1) Medicine (Biomedical
Genetics), Boston University, Boston, MA; 2) Biostatistics, Boston University,
Boston, MA; 3) Ophthalmology, Boston University, Boston, MA; 4) Neurology, Boston University, Boston, MA; 5) Epidemiology, Boston University,
Boston, MA; 6) Neurology, Columbia University, New York, NY; 7) Molecular
Physiology and Biophysics, Vanderbilt University, Nashville, TN; 8) The
John P. Hussman Institute for Human Genomics, University of Miami, Miami,
FL; 9) Molecular Genetics, Niigata University, Niigata, Japan; 10) Pathology
and Laboratory Medicine, University of Pennsylvania School of Medicine,
Philadelphia, PA.
The Alzheimer Disease Genetics Consortium (ADGC) has performed
genome-wide association (GWA) studies for late-onset Alzheimer’s disease
(LOAD) in European American (EA), African American (AA), and Japanese
(JPN) populations. GWA analyses were conducted separately in samples
containing 22,289 EA (11,641 cases, 11,341 controls), 5,028 AA (1,489
cases, 3,539 controls), and 1,845 JPN (951 cases, 894 controls) subjects
using imputed genotypes for 7,911,420 SNPs in a logistic regression model
adjusting for age and sex. Results from the three groups for the total sample
(ALL) and subgroups of APOE ε4-positive(58% of cases and 27% of controls)
and ε4-negative (38% cases and 69% controls) subjects were combined
using meta-analysis. Top-ranked SNPs in transethnic meta-analysis were
further examined for ethnic-specific or transethnic signals by evaluating
results within each ethnic group. Significant associations (transethnic meta
P-value <10 −6) were identified with SNPs in novel LOAD loci in the ALL
(MTOR, HBEGF, CASS4, and near ODZ2) and in ε4-positive (METTL19,
GRIK2, GJC1, TRMT44, SH2D4B, and near EOMES) groups. We observed
a novel genome-wide significant (GWS) association (P<5×10 −8) with a GJC1
SNP in the ε4-positive group, but this finding was evident in Caucasians
only (rs11871429: PEA=9.1×10 −9, PAA=0.83, PJPN=0.17, and PEA-AA-JPN=
6.1×10 −8). Associations for other novel loci were most significant in the
transethnic meta-analysis (top result: SH2D4B SNP rs12772279 in ε4-positive subjects; PEA=4.0×10 −5, PAA=0.015, PJPN=0.16, PEA-AA-JPN=2.8×10 −7).
Top-ranked transethnic signals showed the same direction of the effect but
the minor allele frequency varied substantially across the ethnic groups. In
addition to genes in the APOE region, GWS results for previously established
LOAD loci were observed with SNPs in BIN1, PICALM, CR1, MS4A6E, and
PTK2B in ALL subjects (best SNP: rs586274 in PICALM; PEA=5.8×10 −11,
PAA=0.633, PJPN=3.0×10 −4, PEA-AA-JPN=1.1×10 −12). Among other established LOAD loci in EA, different SNPs showed the smallest P values in the
AA or JPN samples, whereas top-ranked SNPs were not significant (P>0.1).
An attempt to replicate and extend the top findings in independent Caucasian
(N~9,000) and Hispanic (N~5000) samples is underway. This study suggests
that the transethnic approach can detect novel LOAD susceptibility genes
whose effects are either independent of or influenced by APOE genotype.


Pathway-based polygenic scores and variation explained by functional SNPs in schizophrenia. K.K. Nicodemus, G. Donohone, D. Tropea, D. Morris, K. Kendler, F. O'Neill, D. Walsh, E. Riley, M. Gill, A. Corvin, Wellcome Trust Case Control Consortium 2. 1) Department of Psychiatry, Trinity College Dublin, Dublin, Ireland; 2) Virginia Commonwealth University, Virginia, USA; 3) Queen’s University Belfast, Belfast, UK; 4) Health Research Board, Dublin, Ireland.

Schizophrenia (SZ) is a complex disorder with aetiology due to multiple genetic effects, which are likely to be enriched in specific pathways. Using experimentally-derived pathways from SZ-associated genes in the cellular adhesion molecule (CAM), DISC1, NRXN1, TCF4, and ZNF804A pathways, we assessed the polygenic score contribution to variation explained in case status in the WTCCC2 schizophrenia case (N = 1378) control (N = 1086) GWAS. Further, we assessed the contribution of functional SNPs (fSNPs) versus intronic SNPs not in LD with fSNPs in amount of variation explained.

We created polygenic scores based on p-value thresholds and weighted by the log10(odds ratio) from the PGC1 SZ GWAS. The number of genes/ SNPs included were: CAM (125/6332), DISC1 (130/5384), mir-137 (474/ 21368), NRXN1 (131/5931), TCF4 (568/9624) and ZNF804A (138/3086). Restricting to fSNPs produced (% total): CAM 4.8, DISC1 5.8, mir-137 5.0, NRXN1 5.3, TCF4 8.9 and ZNF804A 8.5. Nagelkerke’s R2 was estimated using logistic regression. P-values were obtained using a LRT. We compared the R2 from the set of functional SNPs to that obtained using 1000 size sets of randomly-selected intronic SNPs not in LD (R2 < 0.2) with the fSNPs.

The amount of variation explained (R2, LRT p) was: CAM (0.25%, 0.031), DISC1 (0.95%, 2.9e-05), TCF4 (0.76%, 0.00018), ZNF804A (0.91%, 4.3e-05), NRXN1 (0.27%, 0.026) and mir-137 (2.7%, 1.1e-12). The use of fSNPs explained a significant proportion of variation estimated (% R2 attributable to functional SNPs): CAM 84, DISC1 54, TCF4 54, ZNF804A 32, mir-137 5.2, and NRXN1 1.2. Further, the % of LE intronic SNP sets exceeding the R2 for the fSNPs was: CAM 1.9%, DISC1 8.2%, TCF4 29%, ZNF804A 0% mir-137 56.0% and NRXN1 92%, indicating fSNPs may explain more variation.

The polygenic score based on fSNPs explained a significant proportion of the total variation explained and was larger than that to be able to be explained using intronic SNPs in LE with fSNPs. Assuming equal variation explained by each SNP, the fSNP polygenic score should explain 4.9% of the total; in all but 2 pathways the percent variation explained by fSNPs was > 30%. The use of fSNPs may improve interpretability of the polygenic score by restricting to SNPs with known function and by reducing the number of SNPs to examine in follow-up studies.

Meta analysis identifies TSNARe1 as novel Schizophrenia / Bipolar susceptibility locus. P. Sleiman1, D. Wang2, J. Gilliesen3, D. Hadley3, R.E. Gur4, N. Cohen4, Q. Li4, H. Hakonarson4, Janssen-CHOP Neuropsychiatric Genomics Working Group. 1) Center Applied Genomics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA; 2) 4Janssen Research & Development, LLC, Raritan, NJ; 3) Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA 19104, USA.

We carried out a meta-analysis of combined schizophrenia, schizoaffective, and bipolar cohorts that resulted in the identification of six genome-wide significant loci, including one novel locus at chr2q24.3, encompassing TSNARe1 (P=1.28x10^-9). The study included 13,394 schizophrenia and bipolar cases and 34,676 controls. Of these, 3,182 schizophrenia (of which 377 were classified as schizoaffective) cases and 1,032 bipolar I cases were collected from 28 clinical trials conducted by Janssen Research & Development, LLC. These samples were matched to 15,277 and 8,000 controls, respectively, from the biorepository at the Center for Applied Genomics (CAG) at the Children’s Hospital of Philadelphia (CHOP). In addition, 1,157 cases meeting DSM-IV-TR criteria for schizophrenia or schizoaffective disorder from the Center for Applied Genomics (CAG) at The Children’s Hospital of Philadelphia and the Department of Psychiatry at the University of Pennsylvania School of Medicine and 2,107 controls from the biorepository at CAG were also included in the analysis. The remaining 8,023 schizophrenia cases and 9,292 controls were part of the Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC), as previously described, and were obtained from the NIMH as schizophrenia distribution 9 (https://www.nimhgenetics.org/). The function of the TSNARe1 gene remains unknown. A recent publication suggests it may have evolved within the vertebrate lineage from the harbinger transposon superfamily. Bioinformatic predictions based on phylogenetic footprinting indicate that TSNARe1 may be conserved and SNARE (soluble N-ethylmaleimide-sensitive factor attached protein receptor) proteins and have SNAP receptor activity. TSNARe1 may therefore have a vertebrate-specific function in intracellular protein transport and synaptic vesicle exocytosis.

Polymorphisms in the CACNA1C gene region and amygdala activation during encoding of negative emotional stimuli. L. Gschwind1, C. Vogler2, D. Cournet1, A. Milnik1, V. Freyt1, K. Spalek1, D. de Quervain1,2,3, A. Papasotropoulos1,2,3, 1) Molecular Neurosciences, University of Basel, Basel, Basel, Switzerland; 2) Cognitive Neurosciences, University of Basel, Basel, Basel City, Switzerland; 3) Psychiatric University Clinic, University of Basel, Basel, Basel City, Switzerland; 4) Department Biozentrum, Life Sciences Training Facility, University of Basel, Basel, Switzerland.

Background: Imaging genetics links brain imaging data (e.g. functional Magnetic Resonance Imaging (fMRI)) to genetic variation. Narrowing down the neuronal correlates of psychopathology, i.e. identifying patterns of altered brain activation could serve as valuable endophenotype in genetic research. Differences in amygdala activation, a brain region that is involved in processing of emotional content, have been frequently reported comparing psychiatric patients to healthy controls. Recent studies reported genotype dependent activation differences of the amygdala for the intragenic SNP rs1006737 in CACNA1C (Wessa et al., 2010 and Tisi et al, 2013), a gene implicated in psychiatric disorders and hypothesized that intragenic variation of CACNA1C might affect amygdala activation also in healthy individuals. Therefore we tested variants in the CACNA1C gene for association with amygdala activation in response to emotional stimulus material during an fMRI session in a large sample of healthy individuals. Methods: A sample of N=917 healthy young Swiss individuals encoded a set of 24 negative emotional pictures taken from the International Affective Picture System in a 3 Tesla fMRI scanner. Genotyping of the CACNA1C gene was done using the Affymetrix Human SNP Array 6.0. A total of 183 SNPs fulfilled genotype quality criteria (MAF > 5%, Deviation from HWE p<0.01). Automated segmentation procedures of structural Magnetic Resonance images were used to create individual masks for the accurate extraction of individual brain activation data. Subsequently, amygdala activation was calculated contrasting negative emotional pictures vs. scrambled pictures. Genetic associations were tested under the assumption of an additive model using the Plink software package. Results: The lowest p-value of the 183 SNPs assessed was found for SNP rs1790239 (p=0.00052), which did not survive correction for multiple testing (Bonferroni corr. p < 0.09). Further studies are warranted to elucidate the role of genetic variation of CACNA1C for amygdala activation in healthy individuals.
1290T Genetic variation underlying amygdala-volume is highly enriched with schizophrenia susceptibility variants in healthy young individuals. P. Lee1,2, A. Holmes2, L. Gallagher3, L. Germain3, M. Hollinshead3, J. Roffman3, R. Buckner1,4, J. Smoller1,5,6,7. 1) Massachusetts General Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Stanford Center for Psychiatric Research, Broad Institute, Cambridge, MA; 4) Harvard University, Cambridge, MA. 

Background: Variations in amygdala volume have been implicated in the pathophysiology of schizophrenia, bipolar disorder, and autism, but the genetic basis of these volumetric differences remains undefined. Here we report genome-wide association analysis of a cohort of 1,426 healthy individuals and the genetic relationship between this brain structure and major neuropsychiatric disorders. 

Methods: Our study is based on the Harvard/MGH Brain Genomics Superstruct Project (GSP), a neuroimaging and genetics study of brain and behavioral phenotypes comprising more than 3,500 healthy subjects. Structural MRI images of amygdala volume were assessed for all GSP subjects using a standardized imaging protocol (3T, 12-channel coil, T1). Genome-wide genotyping was performed on 1,870 subjects of European ancestry, yielding the discovery GWAS data of 1,140,419 SNPs. All study subjects were young adults with no history of psychiatric illnesses or major health problems (18<age<35). Imputation using the 1,000 Genomes produced the allele dosages of 8,393,342 SNPs (R2>0.3; MAF>0.01), for which single-variant association was assessed using linear-regression. Covariates included age, gender, handedness, intracranial volume, scanner, console, and MDS factors to control for population stratification. Genetic relationships of amygdala-volume-associated SNPs to major neuropsychiatric disorders were examined using multivariate enrichment analyses. 

Results: Using a neuroimaging GWAS analysis of healthy young adults, we identified genetic variants influencing individual differences in amygdala volume. We also found a potential role for schizophrenia susceptibility variants in modulating normal variation of amygdala volume, suggesting an etiologic link between amygdala structural changes and emotional/cognitive abnormalities present in this serious brain disorder.

1292W SZT2 mutations in infantile encephalopathy with epilepsy and callosal dysgenesis. L. Basel-Vanagaite1,2, J. Takanashi1,2, J. Barkovich1,2,4, J. Reich7, L. Korenreich2, H. Thiele10, H. Bode9, I. Lagovskys2, D. Dahary1, H. Haviv1,5, M. Pasmanik-Chor12, P. Nürnberg10, C. Kubisch6, M. Shohat1,2,3, A. Macaya7, G. Borck1, Rabin Medical Center, Petah Tikva, Israel; 2) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Feilsenstein Medical Research Center, Petah Tikva, Israel; 4) Schneider Children’s Medical Center of Israel, Petah Tikva, Israel; 5) Rambam Health Care Campus, Haifa, Israel; 6) Assaf Harofeh Medical Center, Zerifin, Israel; 7) Vell d'Hebron Research Institute (VHIR), Autonomous University of Barcelona, Barcelona, Spain; 8) Institute of Human Genetics, University of Ulm, Ulm, Germany; 9) Children’s Hospital, University of Ulm, Ulm, Germany; 10) University of Cologne, Cologne, Germany; 11) Toldot Genetics Ltd., Hod Hasharon, Israel; 12) Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel.

Epilepticencephalopathies are severe developmental disorders characterized by early onset epilepsy and poor neurological outcome. The genetic origins and underlying molecular processes of early onset epileptic encephalopathies have been described in only a few well defined syndromes. The objective of this study was to identify the disease-related gene in two unrelated patients with a distinctive phenotype characterized by refractory epilepsy, absent developmental milestones, thick corpus callosum and cavum septum pellucidum on brain MRI. In order to identify the disease-associated gene, we performed whole exome sequencing of the DNA of two probands, a 10 year-old girl of Iraqi Jewish origin and a 9 year-old Spanish boy who both presented with an infantile encephalopathy with refractory epilepsy. In both cases, we identified biallelic mutations of the Seizure Threshold 2 (SZT2) gene. Both patients presented with a severe infantile encephalopathy with intractable seizures, thick corpus callosum and cavum septum pellucidum. While the cellular function of SZT2 is unknown and no diseases caused by SZT2 mutations have been reported in humans, SzT2 is known to influence seizure threshold and epileptogenesis in mice. Truncating mutations of SzT2 can confer low seizure thresholds and embryonic lethality in mice, phenotypes that are well consistent with the severe seizures and developmental arrest in our patients.
cにあるアリズメの遺伝的リスクを指す、およびヒポキャンパルおよびアミグダラの体積が、これらはどの程度をAD病の進展に影響を与えるかを示している。意識の早期影響を増加させることは、これらの遺伝的変異が作用することを示唆するものである。さらに、識別可能なバイオマーカーが存在し、何らかの形で影響を及ぼすことが示唆される。

1294F Genome-wide linkage analysis suggests oligogenic regulation of the human parieto-occipital 10-Hz rhythmic activity. E. Saimela1, H. Renvall2, M. Vihtel3, J. Kujala3, R. Saimelin3, J. Kero1, 2

The human cerebral cortex shows several intrinsic oscillations (rhythms) that can be characterized with non-invasive imaging methods such as magnetoencephalography (MEG). The most prominent of them is the 10-Hz rhythm recorded over the parieto-occipital cortices. This rhythm is strongly attenuated for example by opening the eyes, and its reactivity is widely used to probe cortical functions in both healthy and clinical populations. However, little is known about its underlying molecular mechanisms. To study the possible genetic determinants of the parieto-occipital 10-Hz rhythm in a normal population, we measured spontaneous brain activity with MEG in 98 pairs of healthy siblings while the subjects had their eyes closed and open (3 minutes each). The cortical activity was recorded with a 306-channel Elekta Neuromag magnetometer, and amplitude spectra at each channel were calculated using Fast Fourier Transformation (FFT).

The subjects were genotyped on Affymetrix 250K Sty1 SNP arrays, yielding genotypes for more than 28,000 single-nucleotide polymorphisms (SNPs) after filtering for genotyping quality and linkage disequilibrium.

The brain activity was quantified from the difference spectra between “eyes closed” and “eyes open” conditions. Width of the main spectral peak at -10Hz, its peak frequency and peak strength were measured at the maximum channels in three regions over the parieto-occipital cortices. The peak strength of the rhythm were highly heritable (h2 = 0.87), and we analyzed the variance component-based analysis revealed linkage both for the strength and the width of the spectral peak. The highest linkage peak was seen for the width of the spectral peak over the left parieto-occipital cortex on chromosome 10p23.2 (LOD = 2.814, p < 0.03 based on 1001 permutations with Merlin). This genomic region contains several functionally plausible genes, including GRIN1 which encodes a subunit of glutamate receptor channels that mediate the fast excitatory synaptic transmission in the central nervous system. Altogether, 12 suggestive or significant linkage peaks were seen across the 9 phenotypes tested; in 1001 permutations, the probability of observing 12 equally high peaks by chance was 0.046. Overall, the results demonstrate the potential of genetic analysis in linking macroscopic cortical phenotypes with molecular-level processes controlled by specific genes.

1295W Genetic resilience to neurodegeneration in the presence of tau pathology. T. Hohman, M. Koran, T. Thornton-Wells. Molecular Physiology and Biophysics, Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

The disease cascade in Alzheimer’s Disease (AD) involves tauopathies, β-amyloid plaques, progressive neurodegeneration, and cognitive impairment. Certain individuals present with pathologic protein biomarkers but no clinical manifestation of AD. The current project sought to identify genetic variants that modify the relationship between biomarkers of pathology and ventricular volume (neurodegeneration). We also investigated whether the observed protein-protein interactions had an effect on levels of neuroinflammatory markers in plasma. We used ventricular volume from Magnetic Resonance Imaging (MRI) images quantified in Freesurfer as our outcome measure. Our statistical model controlled for age, education, diagnosis, and gender. Our term of interest was a gene x Phosphorylated Tau (Ptau) interaction. Ptau was quantified from protein levels in cerebrospinal fluid (CSF). Genes were identified using a Genome-wide Association Study (GWAS) approach to look for significant single nucleotide polymorphism (SNP) x Ptau interactions. We corrected for multiple comparisons using a False Discovery Rate threshold of p < 0.05. In addition, we tested for consistency across two subsets of the ADNI study. We identified one SNP x Ptau interaction that was consistent across both the ADNI-1 and ADNI-2/GO data sources: POT1 (rs4728029; T = 5.269; p = 1.52x 10-7). The POT1 x Ptau interaction explained two percent of the variance in ventricular volume even when controlling APOE and the other covariates previously mentioned in the model. POT1 expression has been related to telomere length in previous research (Kondo et al., 2004), and telomere length has been associated with a neuroinflammatory response to Alzheimer’s disease (Panossian et al., 2003). Therefore, in posthoc analyses we tested the Ptau x rs4728029 interaction in relation to neuroinflammatory markers measured in plasma. Indeed, this interaction was related to levels of interleukin 6 receptor (IL6R), providing additional support for a possible neuroinflammatory mechanism that explains the observed genetic interaction. These results suggest that genetic variation may modulate the neuroinflammatory response to Alzheimer’s tau pathology and ultimately modify risk for the neurodegenerative cascade. Future work will use a more focused pathway approach to further investigate genetic resilience related to a beneficial neuroinflammatory response.
1296T  Genetic interactions found between calcium channel genes modulate amyloid load measured by positron emission tomography. M. Koran, T. Hohman, T. Thornton-Wellis, Alzheimer's Disease Neuroimaging Initiative, Vanderbilt University, Nashville, TN.

Late-onset Alzheimer's disease (LOAD) is known to have a complex, oligogenic etiology, with considerable genetic heterogeneity. We investigated the influence of genetic interactions between genes in the Alzheimer's disease (AD) pathway on amyloid-beta (Aβ) deposition as measured by AV45 ligand positron emission tomography (PET) to aid in understanding this disease's genetic etiology. The Alzheimer's Disease Neuroimaging Initiative (ADNI) cohorts were used for discovery and replication. We discovered and validated a significant genetic interaction between the genes RYR3 and CACNA1C, both of which encode calcium channels expressed in the brain. The results shown here support previous animal studies implicating interactions between these calcium channels in amyloidogenesis and suggest that the pathological cascade of this disease may be modified by interactions in the amyloid-calcium axis. Future work focusing on the mechanisms of such relationships may inform targets for clinical intervention.

1297F  Genetic interactions within inositol-related pathways are associated with longitudinal changes in ventricle size. T. Thornton-Wellis1, T. Hohman1, S. Meda2, M. Koran3. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Olin Neuropsychiatry Research Center, Harford Hospital, Hartford, CT.

The genetic etiology of late onset Alzheimer disease (LOAD) has proven complex, involving clinical and genetic heterogeneity and gene-gene interactions. Recent genome wide association studies (GWAS) in LOAD have led to the discovery of novel genetic risk factors; however, the investigation of gene-gene interactions has been limited. Conventional genetic studies often use binary disease status as the primary phenotype, but for complex brain-based diseases, neuroimaging data can serve as quantitative endophenotypes that correlate with disease status and more closely reflect pathological changes. In the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, we tested for association of genetic interactions with longitudinal MRI measurements of the inferior lateral ventricles (ILVs), which have repeatedly been shown to be associated with LOAD status and progression. We performed linear regression to evaluate the ability of pathway-derived SNP-SNP pairs to explain variance in the ILVs and the pons. Both genes belong to the inositol phosphate signaling pathway which has been previously associated with neurodegeneration in AD.

1298W  Associations of the rs4818 polymorphism in the COMT gene with demyelination of the inferior longitudinal fasciculus in the white matter of the brain and cognition in schizophrenia patients. Z. Kikinis1, K. Green1, M. Giwerc1, S. Bouix2, N. Makris2, N. Schreiber2, G. Corfas2, R. Kuchertapatil3, R. Kikinis1, M. Kubicki1, M. E. Shenton1, 2. 1) Harvard Medical School, Boston, MA; 2) Harvard Medical School, Brockton, MA.

A single nucleotide polymorphism (SNP, rs4818) located at the 4th intron of the Catechol-O-methyltransferase (COMT) gene, has been observed to cause changes in expression of the COMT protein. COMT is a schizophrenia susceptibility gene and variants of rs4818 have been reported to be associated with variations in cognitive performance in schizophrenia. However, the impact of rs4818 polymorphism on the white matter of the brain has yet to be investigated. Here, we compared CC, CG and GG carriers of the rs4818 SNP for changes in the inferior longitudinal fasciculus (ILF). The ILF is a long association fiber tract connecting occipital and temporal lobes in the white matter of the brain. We used MR-Diffusion Tensor Imaging (MR- DTI), to analyze white matter microstructure of this tract and we applied neuropsychological tests to assess cognitive performance. Methods: MR-DTI images were acquired on 30 patients with chronic schizophrenia and 26 matched healthy controls and tractography of ILF was performed. Changes in MR-DTI output measures, such as Fractional Anisotropy (FA), trace, Axial Diffusivity (AD) and Radial Diffusivity (RD), were analyzed as they might reveal microstructural changes of axons, such as changes in myelination. DNA was extracted from saliva. The effect of the rs4818 polymorphism on cognition was tested using scores of Wisconsin Card Sorting test (WCST). Results: Schizophrenia patients with the CC genotype (N=10) at rs4818 exhibited higher values in RD, higher values in trace and lower correlation between AD and FA in the ILF in the right hemisphere, which were all statistically significant (p=0.008, p=0.02, p=0.05). Increases in RD, while no changes in AD and decreases in FA have been reported previously in animal studies as consequence of demyelination in brain white matter. Schizophrenia patients with CC/GG and the control groups with either genotype did not differ in the DTI values. Further, RD significantly and positively correlated to the scores of several subtests of the WCST (p<0.02) in the group of schizophrenia patients with the CC genotype at rs4818. This suggests that demyelination of the axon is associated with poorer performance on the test. Conclusion: Our results suggest that the CC genotype at rs4818 of the COMT gene is associated with demyelination of the white matter tract ILF in schizophrenia patients. These changes in myelination might affect working memory and cognition in schizophrenia patients with this particular genetic make up.

1299T  Identification of robust biomarkers of neuronal and glial metabolic changes in spinocerebellar ataxia type 1,2,3 and 7. T. Mawusi1, D. Mawusi2, T. M. Nguyen1, M. E. Shenton1, M. Bouix3, A. K. Hauser4, J. M. Delisle4, D. Chedotal3, R. Kucherla5, M. Klump6, T. Kikinis1, K. Green1, G. Corfas2, R. Kuchertapatil3, M. E. Shenton1, 2. 1) Department of Genetics, Hôpital de La Salpêtrière, Paris, France; 2) CENIR, Institut du Cerveau et de la Moelle, Paris, France; 3) University of Minnesota, Center for Magnetic Resonance Research, Minneapolis, Minnesota, USA.

Background: Spinocerebellar ataxias (SCAs) belong to the group of polyglutamine repeat disorders and lead primarily to neurodegeneration in the cerebellum and the pons. We recently demonstrated that even the most sensitive clinical scores would require large number of patients to assess any therapeutic benefit. Therefore, the identification of robust biomarkers is critical to assess disease progression for therapeutic development. Methods: 1H-NMR spectroscopy was performed at 3T to determine the neurochemical profile of 24 metabolite concentrations in the vermis and pons of a unique cohort of 66 SCA patients - SCA1 (N=18), SCA2 (N=13), SCA3 (N=22) and SCA7 (N=13) - as well as in healthy controls with similar median age (N=34). Results/Interpretation: Compared to controls, SCAs patients displayed a significant decrease of neuronal metabolites, N-acetylaspartate and glutamate, but increased glia-related metabolites, glutamine and myoinositol. The neuronal loss in both affected regions was associated with a significant increase in creatine and phosphocreatine suggesting compensatory energetic mechanisms. Of note, there was a strong negative correlation between ataxia rating score (SARA) and total N-acetylaspartate in the pons of SCA2 (r = -0.7942, p = 0.0186), SCA3 (r = -0.7868, p= 0.0008) and SCA7 (r = -0.7614, p = 0.0105). In the vermis, SARA score correlated with total creatine in SCA1 (r = -0.7073, p = 0.0307), SCA2 (r = -0.6327, p = 0.0253), SCA3 (r = -0.6759, p = 0.0072). A correlation with myoinositol was also found in SCA2 (r = 0.6142, p = 0.0336) and SCA3 (r = 0.5831, p=0.0055). The PCA confirm that neuronal metabolites (N-acetylaspartate and glutamate) in the vermis and the pons vary inversely to glia-related metabolites (myoinositol) and energy-related metabolites (creatine and phosphocreatine), suggesting compensatory mechanisms to the neuronal loss.
1300F Analysis of Structural Variants in the DISC1 gene: Association Test with P300 ERP and Brain Volumes. V. De Luca1, G. Spalletta2. 1) Dept Neurogenetics. Univ Toronto, Toronto, ON, Canada; 2) Neuropsychiatry. IRCCS Santa Lucia Foundation, Italy.

In a large Scottish family with a (1;11)(q42;q14.3) translocation, the carriers of this balanced translocation were significantly associated to reduction in the amplitude of the P300 event-related potential (ERP). In healthy controls, we tested the association between two non-synonymous SNPs in DISC1 gene with P300 amplitude and the volumes of the brain structures that are the putative P300 generators. We have investigated the measurement of the visual P300 event-related potential (ERP) during a modified Stroop task. For this study, 32 healthy subjects (drug free) were recruited for the ERP arm and 100 subjects were recruited for the structural MRI arm. The mean age in the ERP was 34.53±8.5. There were 18 males and 14 females. The analysis of the SNP rs821616 (Cys704Ser) was performed using the AB7700. For the neutral condition, there was no difference between the subjects who were carrying the variant 704Ser and those who did not carry this variant [F (1/ 26) = 1.732, p=0.200]. In the structural MRI arm, we did not find association with the SNP rs821616. Although our results did not show an association between Cys704Ser in DISC1 and P300 amplitude, this pilot study suggests a new methodology to combine multiple samples with related endophenotypes to test the pleiotropic effect of specific functional variants.

1301W Clinical and Molecular Investigation of Spinocerebellar Ataxia with Hypogonadism: Not So Far, Not So Close. C. Lourenço1, C. Sobreira1, P. Frassinette de Medeiros2, M. Gonzalez3, F. Speziani3, S. Zuchner4, W. Marques Jr1. 1) Neurology, Univ Sao Paulo, Ribeirao Preto, SAO PAULO, Brazil; 2) Genetics Service , Federal University of Campina Grande, Campina Grande, Brazil; 3) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

BACKGROUND: The association between cerebellar ataxia and hypogonadism was first described in four sibs by Holmes in the beginning of the XX century, and has since become known as Holmes type ataxia. Several syndromes with hypo/hypergonadotropic hypogonadism and ataxia have been published, however there is a remarkable clinical heterogeneity among them. Here, we present the clinical data and molecular/biochemical studies of nineteen Brazilian patients with cerebellar ataxia and hypogonadism. MATERIAL AND METHODS: All patients were evaluated in the neurogenetics clinics by geneticists, neurologists and endocrinologists. Brain MRI, ophthalmological exam, EMG/NCV, hormone and biochemical tests, screening for CDG disorders (IEF and MS analysis), karyotype, muscle biopsy with chain respiratory enzyme assays and measurement of coenzyme Q10, molecular tests for Friedreich ataxia and for SCAs (types 1, 2, 3, 6 and 7) were performed in the course of the investigation. Exome sequencing was performed in selected patients with a consistent phenotype. RESULTS: All patients had cerebellar ataxia, but the age of the onset was variable; it was worthy to note that ten patients had early-onset ataxia. (in the first decade of life). Consanguinity of parents was noted in three families; seven patients had hypergonadotropic hypogonadism. Mental retardation was seen in five patients with hypergonadotropic hypogonadism. None of the patients had chromosomal anomalies. Molecular tests for Friedreich and SCAs 1, 2, 3, 6 and 7 were all negative. Optic atrophy and retinalchoroidal degeneration were found in five patients; axonal neuropathy was present in four patients. Cerebellar atrophy with pons or prominent vermis involvement was a constant feature. In two patients with ataxia and hypergonadotropic hypogonadism, coenzyme Q10 deficiency was confirmed in muscle biopsy. Two unrelated adult patients with hypergonadotropic hypogonadism had biochemical features of CDG Ia. One family - with four affected sibs - have features consistent with a rare neurological disorder, Boucher-Neuhauer syndrome. CONCLUSIONS: The association between cerebellar ataxia and hypogonadism comprise heterogeneous entities whose clinical investigation can enlighten the pathological basis of these fascinating neuroendocrinological syndromes. Screening for CDG and Coq10 deficiency should be done in such patients as a part of the work-up investigation.

1302T Novel ATM mutation in late-onset autosomal recessive cerebellar ataxia with neuropathy. H. Shimazaki1, R. Sugaya2, J. Honda1, A. Meguro1, I. Nakano2. 1) Neurology, Jichi Medical University, Shimotsuke, Tochigi, Japan; 2) Hematology, Jichi Medical University, Shimotsuke, Tochigi, Japan.

Novel ATM mutation in late-onset autosomal recessive cerebellar ataxia with neuropathy (ARSCA) comprises many types of diseases. Most frequent ARSCA is Friedreich ataxia, but other types are relatively rare. We encountered a consanguineous family with two patients of late-onset cerebellar ataxia with neuropathy. We attempt to identify the causative gene mutation of this family with ARSCA. Methods: We investigated the proband with detailed neurological examination, blood examination, brain MRI, electrophysiological study, nerve biopsy and gene analyses. Results: The neurological examination revealed cerebellar ataxia, hand tremor, distal muscle wasting and diminished tendon reflexes. We had no conjunctival telangiectasias and showed immunodeficiency. Blood examination showed slightly elevated AFP. Brain MRI demonstrated marked cerebellar atrophy. Electrophysiologic study and nerve biopsy showed axonal neuropathy. We could not detect GAA repeat expansions of FXN gene, and sequencing analyses could not reveal pathologic substitutions in the APTX, SETX, and TDP1 genes. Whole-exome sequencing (WES) could identify the novel homozygous missense mutation in the ATM gene. This homozygous mutation was found in another patient, co-segregated within the family members and not found in 200 Japanese control DNAs. Subsequently, the patients suffered and died from bile duct cancer and chronic lymphocytic leukemia. Conclusion: We could identify ATM mutations in adult, late-onset ARSCA family. Ataxia-Telangiectasia (AT) is usually early-onset and show immunodeficiency and telangiectasia. We should consider AT even in late-onset ARSCA without telangiectasia and immunodeficiency. WES is one of the useful methods for identify the causative mutation in atypical adult-onset autosomal recessive cerebellar ataxias.

1303F Modelling collybistin deletion neuropathology in induced pluripotent stem cells (iPSC)-derived neurons. A. Sarte1,2, K.G. Oliveira1, C.O.F. Machado1,2,3, C. Rosenberg4, F. Kok1,3, M.R. Passos-Bueno1. 1) Genetics and Evolutionary Biology Department, Institute of Biosciences, University of Sao Paulo, Brazil; 2) Experimental Research Center, Albert Einstein Jewish Hospital, Sao Paulo, Brazil; 3) Clinics Hospital, School of Medicine, University of Sao Paulo, Brazil; 4) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

Corrybistin (CB) is a neuron-specific guanine nucleotide exchange factor implicated in inhibitory synapse development and plasticity that cluster and localize gephyrin and inhibitory neurotransmitter receptors to the postsynaptic membrane. We have identified a Brazilian patient with a deletion of the entire collybistin gene (ARHGEF9, Xq11.2) who shows severe mental retardation, epilepsy, and autistic behavior. In order to gain further insight into the underlying cellular neuropathology, we generated iPSC-derived neural progenitor cells (NPCs) and neurons from this patient and two unaffected individuals. iPSCs generated from skin fibroblasts using standard retroviral reprogramming technology. iPSC differentiation into NPCs and neurons was performed under conditions that favor the generation of either excitatory or inhibitory neurons. Expression of pluripotency-, NPC- and neuron-specific markers were analyzed by immunofluorescence and western blotting. mTOR signaling pathway analysis was performed in NPCs treated nor not with rapamycin by western blotting. We observed that all control and patient iPSC clones express pluripotent markers such as Nanog, Oct4, Sox2 and Lin28. NPCs were positive for early neural precursor markers, such as Nestin and Musashi1. Mature neurons were positive for neural markers β-Ill Tubulin and Map2; in addition, these cells also express PSD-95 and GABAR-2 proteins, markers for excitatory and inhibitory synapses respectively. Interestingly, cells treated with rapamycin, which is not detected in normal fibroblasts, was reactivated in control iPSC, NPCs and mature neurons. We did not detect any significant differences between patient and control cells with respect to the reprogramming and differentiation capacities. However, iPSC and NPCs derived from the patient showed increased proliferation rate. Because mTOR signaling has been linked to cell proliferation and the CB partner gephyrin interacts with mTOR, we investigated the mTOR signaling pathway in patient-derived NPCs. Compared to control NPCs, our preliminary results suggested altered mTOR signaling in patient NPCs. Together, our results suggest that iPSCs and neurons from the patient carrying a deletion of the CB gene provide very promising model systems to explore the roles of CB in human neural development and physiology, as well as the mechanisms underling the cognitive impairments in CB-deficient patients. Support: FAPESP, CNPq, Autism & Realidade.
1304W  
Genetic basis of Math genetic difficulty. M.R.S. Carvalho1, 2, M.de Miranda1, 3, G. Pena4, A. Julio-Costa4, P. Pinheiro-Chagas3, 7, L. Salvador5, M. Andreata1, V.G. Haase2, 4, 5 1) Departamento de Biol Geral, Univ Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 2) Departamento de Psicologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 3) Pós-Graduação em Genética, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 4) Pós-Graduação em Neurociências, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 5) Pós-Graduação em Saúde da Criança e do Adolescente, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 6) Pós-Graduação em Enfermagem, Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 7) Inserm-CEA Cognitive Neuroimaging unit NeuroSpin Center.

Math learning difficulty (MD) is a common cognitive condition affecting from 3 to 6% of the school-age population. Impacts of MD are low scholarly achievement, low self-esteem, low employability, and low income. Social and economical factors and low educational standards may contribute to MD. However, a primary, genetic predisposition was also proposed. Familial aggregation has been reported for MD, but there is only two reports in the literature analyzing family history. Heritability and relative risks were never estimated. This project was approved by the Ethics Committee of the Universidade Federal de Minas Gerais. It is part of a population based case-control study, in which a random sample of the school age population of Belo Horizonte city, Brazil, was collected. In order to ascertain individuals with MD, a two phase strategy was adopted. In the first phase, a Math achievement test (TDE) was responded by 1564 children. Students with results below the percentile 25 in TDE were further evaluated with a battery of tests covering intelligence, attention, mental state, memory, working memory, fine motor coordination, language, somato-sensorial function. Students with IQ<85 were excluded. A family history was collected with an specifically developed, structured questionnaire composed by simple questions such as ‘can he/she add?’, ‘can he/she subtract?’, ‘can he/she multiply?’, ‘can he/she divide?’, ‘can he/she exchange money’, ‘can he/she discriminate left and right’, ‘can he/she tell the time on the clock?’, ‘can he/she go around by himself/herself’, ‘can he/she write?’, ‘can he/she read?’. Besides, data were collected on health conditions, consanguinity, scholarship, profession, and familial income. Fifty-one historically studied families were selected, as well as 27% of the mothers, 16% of the fathers, and 40% of the brothers and sisters of the affected children. MD frequency among the 1564 students that took part in the first phase was estimated in 7%. No significative differences in sexual proportion were observed. Genetic counseling was offered for MD index cases. Familial recurrence was reported in 32/34 families of the affected children. MD was reported in 27% of the mothers, 16% of the fathers, and 40% of the brothers and sisters of the affected children. These results are similar to those found in the two previous reports. Heritability was estimated in 70% for MD. Both autosomal dominant and X-linked recessive inheritance were detected among the MD families. Support: CNPq, FAPEMIG.

1305T  
Whole Transcriptome Analysis of Fetal and Adult Ts1Cje Brains Provides New Perspectives for Prenatal and Postnatal Treatment of Neurocognition in Down Syndrome. F. Guedj1, J.L.A Pennings2, L.C Graham3, E.L. Newman4, K.A Micske5, D.W. Bianchi6 1) Mother Infant Research Institute, Tufts Medical Center, Boston, MA, USA; 2) Laboratory for Health Protection Research (GBO), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands; 3) Department of Psychology, Tufts University, Boston, MA, USA.

Background: Learning deficits in Down syndrome (DS) emerge in early childhood and are associated with microcephaly and reduced neurogenesis. Although these anomalies originate during fetal life, pre-clinical trials are mainly focused on adults. We have proposed using non-invasive prenatal testing (NIPT) as an opportunity for prenatal treatment to improve cognition in DS (PrenatDiagn 2013;33:614). We hypothesize that prenatal treatment of brain and cognitive anomalies can be achieved using a systems biology approach. As baseline information, we analyzed whole transcriptome changes in embryonic and adult brains of the Ts1Cje mouse model of DS, as well as neonatal behavior via the ultrasound vocalization (USV).

Methods: For gene expression studies, total RNA was prepared from adult cortex and hippocampus (Ts1Cje=6/WT=5) and E15 whole brains (Ts1Cje=5/WT=5) for hybridization to Affymetrix mouse gene 1.0ST arrays. Data were analyzed to identify genes that are differentially regulated. Functional analyses were performed using GSEA and DAVID to identify altered signaling pathways in Ts1Cje brains. For behavioral studies, P7 pups were separated from their dams and the number of USVs was recorded. Motor activity was assessed by estimating the number of grid crossings and rollings on the testing platform.

Results: More genes were significantly differentially regulated in the Ts1Cje vs. WT E15 brains (n=71) compared to the hippocampus (n=30) and cortex (n=7) at FDR-BH < 20%. Interestingly, similar regulation patterns are observed at both stages. Functional analyses highlighted imbalances of several cellular pathways (inflammation, oxidative stress, apoptosis and G-protein signaling) at both stages. Importantly, cell cycle genes (28 genes) and amino acid transporters (7 genes) are exclusively affected in fetal brain. Ts1Cje neonates produced more ultrasonic vocalizations (sum of ranks=274.50) than WT pups (sum of ranks=221.50) (p<0.01, Mann-Whitney test). They also display abnormal motor activity indicated by reduced number of grid crossings (Sum of ranks=141.50) versus WT (Sum of ranks=141.50) (p<0.01).

Conclusions: Data acquired from analysis of the brain transcriptome and early neonatal behavioural screening in Ts1Cje mice suggest novel pathological and behavioral abnormalities that are already present during fetal and neonatal life in DS. These differences provide baseline information on which to evaluate the effect of in utero treatment strategies.
Muzny

Gene expression is enabling a molecular diagnosis. Genomics is Genomic approaches are allowing the identification of novel variants in observed in patients with VRK1 mutations. Patients with rare diseases and axonal neuropathy in mice. We hypothesize that a similar mechanism ride p53 and has crucial roles throughout the cell cycle. It has been proposed that VRK1 acts downstream of CLP1, and the disruption of the latter was shown recently to result in abnormal neuronal apoptosis leading to a similar neurodegenerative phenotype defined by microcephaly, brain dysgenesis and axonal neuropathy in mice. We hypothesize that a similar mechanism of abnormal natural apoptosis may explain the neurologic phenotype observed in patients with VRK1 mutations. Patients with rare diseases and complex clinical presentations represent a challenge for clinical diagnostics. Genomic approaches are allowing the identification of novel variants in genes found in rare disorders enabling a molecular diagnosis. Genomics is also revealing a phenotypic expansion whereby the full spectrum of clinical expression conveyed by mutant alleles at a locus can be better appreciated.

1306F
Monamine Oxidase Deficiency: The clinical relevance of personal genomics in a new developmental brain dysfunction disorder, D. Mornero-De-Luca1, E.R. Riggs2, D.H. Ledbetter3, C.L. Martin4, J.F. Cubells5, 1) Department of Psychiatry Yale University 300 George Street, Suite 901 New Haven, CT 06511; 2) Geisinger Health System 100 North Academy Avenue Danville, PA 17822; 3) Department of Human Genetics Emory University School of Medicine Woodruff Memorial Building, Suite 7213 1638 Pierce Drive Atlanta, GA 30322.

Now considered the standard of care for developmental disabilities, chromosomal microarray testing has opened phenotypic delineation of new neurodevelopmental syndromes, defined by specific molecular criteria. Here we describe the phenotypic, familial, and genetic characterization of a male patient with a deletion involving the X-linked MAOA and MAOB genes, essential for normal neurotransmitter homeostasis. The male proband presented at the age of 27 for psychiatric care. Facial dysmorphology, autism and ID, plus a family history remarkable for four maternal uncles with ID and premature deaths due to stroke, prompted evaluation by array comparative genomic hybridization (aCGH). We used the International Collaboration for Clinical Genomics (ICCG, formerly ISCA) consensus microarray design, containing 180K oligonucleotide probes across the whole genome. To fine-map initial results, we used a custom-designed high-density exon targeted microarray based on the same platform. aCGH revealed a 897 kb deletion on the X chromosome (chrX:42,625,499-43,523,142) that entirely removes MAOA and disrupts the coding region of MAOB. The patient’s phenotype including ID, complete lack of productive language and very limited receptive language, complex wringing limb and body movements, constant motor restlessness, and frequent brief syncopal episodes, is remarkably similar to other individuals with deletions involving MAOA and MAOB, but strikingly different from Brunner syndrome, arising from a point mutation in MAOA alone. The family was advised to refrain from giving the patient foods containing tryptamine, and they were provided a list of sympathomimetic medications to avoid. The clear genotype-phenotype correlation in this individual, and the dramatic family history of premature death, is similar to the presentations of other males with deletions disrupting both MAO-encoding genes. Our observations and prior reports establish MAO deficiency as a new developmental brain dysfunction disorder with explicit clinical implications, namely the need for a low tyramine diet and clear contraindication of sympathomimetic agents. Knowledge of the genetic etiology of the phenotype in these patients allows for a tailored clinical management targeting directly the deficient pathway and serves as an example of personalized medicine.

1307W

Here, we report three children from two unrelated families with a previously uncharacterized complex axonal motor and sensory neuropathy accompanied by severe non-progressive microcephaly and cerebral dysgenesis. We performed whole-genome and targeted whole-exome sequencing in these three affected subjects. Using genome-wide sequence analysis, we identified compound heterozygous mutations in two affected siblings from one family and a homozgyous nonsense mutation in the third unrelated patient in the vaccinia-related kinase 1 (VRK1) gene. VRK1 encodes a serine-threonine kinase that is crucial for cell cycle progression and cell division and is proposed to be involved in nervous system development and maintenance. It is an early response gene that directly phosphorylates and regulates p53 and has crucial roles throughout the cell cycle. It has been proposed that VRK1 acts downstream of CLP1, and the disruption of the latter was shown recently to result in abnormal neuronal apoptosis leading to a similar neurodegenerative phenotype defined by microcephaly, brain dysgenesis and axonal neuropathy in mice. We hypothesize that a similar mechanism of abnormal natural apoptosis may explain the neurologic phenotype observed in patients with VRK1 mutations. Patients with rare diseases and complex clinical presentations represent a challenge for clinical diagnostics. Genomic approaches are allowing the identification of novel variants in genes found in rare disorders enabling a molecular diagnosis. Genomics is also revealing a phenotypic expansion whereby the full spectrum of clinical expression conveyed by mutant alleles at a locus can be better appreciated.

1308T
Evidence for phenotypically-linked bounds on the role of de novo loss of function variation in autism spectrum disorders, E.B. Robinson1, 2, 3, 4, K.E. Samocha1, 2, 3, 4, J. Kosmicki1, 2, 3, 4, J. L. McGrath1, 3, 4, B. M. Neale1, 2, 3, 4, M. J. Daly1, 2, 3, 4, 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Medical and Population Genetics Program, Broad Institute of Harvard and MIT, Cambridge, MA; 4) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA; 5) Center for Biomedical Informatics, Harvard Medical School, Boston, MA; 6) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA.

Background The distribution of cognitive ability in individuals with autism spectrum disorders (ASDs) ranges from severe impairment to high intelligence. While ASD with intellectual disability has been associated with a greater frequency of deleterious genetic events than ASD alone (e.g., Girirajan et al., 2012), the extent to which cognitive variation can be used to delineate bounds on the influence of de novo variation in ASDs has not been well characterized. The goal of this study was to determine whether de novo loss of function (LOF) mutations were associated with ASDs across the range of cognitive ability in 355 sequenced trios from the Autism Consortium and Simons Simplex Collection. Method Participants were included if they attempted any one of the 8 IQ tests used across the consortium that offered a comparable estimate of full scale IQ (e.g., the Differential Ability Scales, the WISC; n=801). To maintain adequately powered subsets, we divided the sample into three groups: 1) measured IQ at or above average (≥100; n=229), 2) measured IQ below average IQ (<100; n=457), and 3) IQ test attempted but not successfully completed (n=115). We compared the observed number of de novo LOF mutations per genome in each group against the expected rate in the general population as estimated in the Exome Sequencing Project (0.09 per exome). Results The ASD sample as a whole featured an excess of de novo LOF mutations (observed rate=0.13) compared to the general population (expected rate=0.09). The association between ASDs and de novo LOF variants became stronger when evaluated separately within the groups with measured IQ <100 (observed rate=0.16, p=2.37e-7) and IQ test attempted but not successfully completed (observed rate=0.21, p=9.72e-5). The difference in de novo LOF rate between the IQ<100 and other groups was also statistically significant (p=0.001). Discussion These findings suggest that de novo LOF variation does not play a role in ASDs when intelligence exceeds the general population average. Further, they suggest that ASDs include etiologic subtypes that can be distinguished through phenotypic variation. We discuss these findings in terms of their relevance to study design in ASD genetics.
Phenotypic categorization of putative pathogenic Copy Number Variants (CNVs) in a population of Autism Spectrum Disorder (ASD) patients. I.C. Conceição1, L. Correia1,2, M. Rama-Coelho1, C. Caffe1, J. Almeida1,2, S. Mouga4, F. Duque4, G. Oliveira4,5, A.M. Vicente1,2,3, 1) Instituto Nacional de Saúde Dr Ricardo Jorge, Lisbon, Portugal; 2) Center for Biodiversity, Functional & Integrative Genomics, Lisbon, Portugal; 3) Instituto Gulbenkian de Ciência, Oeiras, Portugal; 4) Unidade Neurodesenvolvimento e Autismo, Centro de Desenvolvimento, Hospital Pediátrico (HP), Centro Hospitalar e Universitário de Coimbra (CHUC), Coimbra, Portugal; 5) Instituto Biomédico de Investigação em Luz e Imagem, Faculdade de Medicina da Universidade de Coimbra, Portugal; 6) Faculdade de Medicina da Universidade de Coimbra, Portugal.

The in-depth characterization of CNVs in patients with ASD is fundamental to improve the distinction between benign and disease-causing structural variants, defining the clinical and genomic attributes with better predictive power for the development of CNV categorization methods for diagnostic purposes. In this study we characterize and categorize CNVs of clinical significance present in a population of 342 ASD patients, genotyped by the Autism Genome Project using SNP arrays. We selected all high confidence genic CNVs that did not overlap more than 20% in sequence with benign CNVs identified in public databases control populations (N=5459), and included CNVs containing syndromic genes/gene regions associated with ASD. A total of 180 genic CNVs were identified, ranging from 5 kb to 3 Mb, 62% of which were deletions. Genic CNVs included one (73% of all genic CNVs) to 25 genes in a single CNV. Most CNVs (86%) were present in a single individual, but 23 common CNVs with a frequency of 1% or higher, distributed in 6 genomic regions, were identified in 26 individuals. Each of these common CNVs were present in 3 to 4 individuals, and encompassed candidate gene/gene regions for ASD, such as 16p13.11, DYX1, PARK2 and VPS13B. Network analysis of the 314 genes encompassed by these CNVs yielded a network including 85 genes, enriched in central nervous system development and regulation of cellular catabolic processes. These CNVs yielded a network including 85 genes, enriched in central nervous system development and regulation of cellular catabolic processes. A significant excess of BAPQs positive for autistic traits in ASD. CNVs were further categorized according to familial correlations assessed for all pairs using the Social Responsiveness Scale (SRS) for parents and Autism Behavior Questionnaire (BAPQ). An estimate of the IQ score of the autistic subjects was 98±23 by Raven and 90±21 by Stanford-Binet testing with 10-14% <70. Thus, the AGRE sample has a higher IQ distribution than typical for fragile X subjects (mean -40±25). Previous prevalence studies of fragile X in autistics ranges from 0 to 16%; with a mean of -4% (Feinstein 98). Our initial 1.9% is similar to a report of 1.6% among 123 unrelated autistic individuals (Bailey 93), but lower than the 13% we found on an earlier multicenter study of 183 individuals (Brown 86). A growing awareness of fragile X syndrome has decreased the probability of fragile X in the autistic population, and due to higher shifts in the IQ distribution for more recent autism cohorts. The overall observed frequency was 8±1742 or 0.5%. This finding still confirms an association of fragile X and autism exists. We tested to see if there is an association of premutation and intermediate alleles in the AGRE probands. Among the 1525 male probands tested, there were 2 with premutation (59 & 64 CGGs) and 12 with intermediate (45-54 CGGs) alleles for an intermediate prevalence of 0.78%. Among the 206 female probands tested there were 2 with premutations (55, 59) and 7 with intermediate alleles. Since females have two alleles, dividing by 2 gives an intermediate allele prevalence of 1.7% in female alleles or an overall intermediate allele prevalence of 0.98%. Our control value was 1.15%. Thus, there was no excess of intermediate or premutation alleles among the AGRE probands. This finding indicates autism is NOT associated with intermediate (45-54) or premutation (55-200) alleles in the AGRE probands.

Genome-wide association study identifies variants for major depression through age at onset stratification. R.A. Power1, K. Tansey1, H. Buttenschon2, H. Lee3, S. Cohen-Woods4, S. Ripke5, N.R. Wray6, P.G.C.-MDD Working Group7, C.M. Lewis3, 1) Social Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, London, United Kingdom; 2) Institute of Clinical Medicine, Aarhus University, Denmark; 3) Queensland Brain Institute, University of Queensland, Australia; 4) University of Adelaide, Australia; 5) Massachusetts General Hospital Simchon Research Center, USA; 6) Consortium.

Major depressive disorder (MDD) is a highly prevalent and disabling disease. Despite moderate heritability and the pooling of large GWAS samples, no replicable genetic variants have been established for MDD. Here we used age at onset (AAO) to dissect the considerable heterogeneity within MDD, and better understand the underlying genetic architecture of this disorder. In the Psychiatric Genomics Consortium’s sample of 9 studies containing 8,920 MDD cases and 9,521 controls, we split cases into 8 subsets based on AAO percentile within their study (i.e. first 8th of earliest onset cases, second 8th of latest onset, etc.). These octiles were analysed systematically against controls to look at both early and late-onset specific risk factors. A replication threshold of p<5E-7 was used, with SNPs replicated in a sample of 6 studies containing 4,922 cases and 15,141 controls. GCTA and polygenic profile scoring were used to examine the heritability of early vs. late onset MDD, and how it overlapped with other psychiatric disorders. Ten independent SNPs met our replication threshold with AAO specific MDD sub-groups, of which two were genome-wide significant in the meta-analysis with the replication sample. These were rs7647854 on chromosome 3 (p=1.0E-11, OR=0.81) and rs7950328 on chromosome 11 (p=1.0E-08, OR=1.35). Further analysis showed that the effect of these SNPs was specific to the 50% latest onset cases, and increased in effect size as AAO increased. With a median AAO at age 27 across samples, ‘late-onset’ here reflects an adult-onset form of MDD rather than one developing in early age. We found no difference in the heritability of early and late-onset MDD, though did find a significantly greater burden of schizophrenia and bipolar risk variants in early-onset cases. In this study we have identified the first genome-wide significant replications for major depression and helped redefine its clinical picture. Our results suggest that adult-onset MDD is a more homogenous disorder with risk variants of larger effect, while adolescent-onset MDD is genetically more similar to both schizophrenia and bipolar disorder. We also highlight how the approach of more homogenous phenotyping can increase power beyond simply increasing sample size.
1312F
Divergent and convergent quantitative dysmorphic phenotypes among neuropsychiatric patients with rare de novo Copy Number Variants. C.K. Deutsch1, F. Momen-Heravi2, R. Francis3, A.T. Hunt4, J.M. Stoler5, L.B. Holmes6, J. Sebat7. 1) Psychobiology Program, Eunice Kennedy Shriver Ctr, Waltham, MA; 2) Harvard Catalyst, Harvard Medical School, Boston, MA; 3) Clinical Genetics, Boston Children’s Hospital, Boston, MA; 4) Medical Genetics Unit, MassGeneral Hospital for Children, Boston, MA; 5) Beyster Center for Molecular Genomics of Neuropsychiatric Disease, University of California San Diego, La Jolla, CA.

Over the last 25 years, quantitative dysmorphology methods have found increasing application in clinical and medical genetics. This has been facilitated by the publication of Farkas’ atlas Anthropometry of the Head and Face. This craniofacial surface measurement system has become a popular standard reference, providing the most extensive normative database extant for direct craniofacial measurement in North America (N=1312). These anthropometric measurements have been used to render the diagnosis of anomalies both objective and reliable. Until recently geneticists have had to rely on cumbersome hard-copy look-up tables for these norms. Through the support of NIH, we have made the use of these norms more user-friendly by creating a computer-based interactive craniofacial measurement database for them. This software, FaceValue, provides tables of normative measurements conditioned on patient demographics. With these tables, the clinical/ medical geneticist configures operational definitions of anomalies, computes their z-scores as continuously distributed variables, and generates reports based on output measures. Here, we have applied FaceValue to the assessment of dysmorphic features in children with rare de novo Copy Number Variants (CNVs), using an anthropometric protocol to document for the first time clusters of specific, quantitatively-defined anomalies for molecular cytogenetic duplications and deletions. We provide worked examples of craniofacial anomalies among specific CNVs, including cases of 1q21.1 duplication, 15q11.2 deletion, and 22q11.2 deletion and duplication. These cases have also been phenotyped with respect to a spectrum of neuropsychiatric diagnosis, including but not confined to autistic disorders. Further, we present worked examples of dysmorphology observed in cases of teratogenesis in the form of prenatal exposure to anticonvulsants. In addition to distinctive combinations of dysmorphic features, among these patients, convergent phenotypes emerged that are shared among multiple cases with autistic disorders. This convergence may signal common underlying forms of maldevelopment contributed by heterogeneous etiologies. Grant support: Simone Foundation SFAI (NeuroACH Study; Sebat J, PI), Interactive Craniofacial Normative Database (R42DE016442; Deutsch CK, PI), Simons Foundation SFARI (Dysmorphology in Autism; Deutsch CK, PI).

1313W
Late-onset Alzheimer disease neuropathology genomic screen identifies novel loci for neuritic plaque and other AD neuropathology features. K.L. Hamilton-Neilson1, C.W. Beecham2, A.C. Nay3, L.-S. Wang4, E.R. Martin5, R. Mayeux6, J.L. Haines6, L.A. Farrer6, G.D. Schellenberg5, M.A. Pericak-Vance7, T.J. Montine8, Alzheimer’s Disease Genetics Consortium. 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Department of Pathology, University of Washington, Seattle, WA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Taub Institute of Research on Alzheimer Disease, Columbia University, New York, NY; 5) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 6) School of Medicine, Boston University, Boston, MA.

Background: Late-onset Alzheimer disease (LOAD) is a highly heritable neurological disease with several known genetic risk loci (APOE, CR1, etc). Most of these loci have small effects on risk (e.g., odds ratios ~1.1-1.2) and, aside from statistical association, no clear connection to LOAD etiology. To further investigate underlying genetic mechanisms of LOAD, we have performed a genome-wide association study (GWAS) of AD neuropathology and related phenotypes, including a neuropathology-confirmed case-control analysis, and analyses of neuropathology features, including neuritic plaques (NP), lewy bodies (LB), amyloid angiopathy (AA), medial temporal sclerosis (MTS), AD Braak stage, and vascular brain injury (VBI). We used this expanded neuropathology approach to limit the effects of phenotypic heterogeneity, and provide additional insights into AD subphenotypes. Methods: We examined 4,914 samples from 11 datasets in the Alzheimer Disease Genetics Consortium which were genotyped with high-density chips and imputed to a 1,000 Genomes Project reference panel. Primary pathology was assessed on ordinal scales (e.g., none, sparse, moderate, or frequent NP); statistical analyses were performed on either the ordinal or binary (i.e., presence or absence) scales as appropriate. Association was performed using logistic regression for binary traits and polytomous logistic regression for ordinal traits, followed by meta-analysis. Subjects examined included 3,887 neuropathologically-confirmed LOAD cases and 1,027 neuropathologically-confirmed cognitive controls. Results: Associations of APOE and BIN1 with LOAD were confirmed. Additionally, several novel LOAD associations were found, including PHF21B (P=2.0×10-8), and SMOX (P=9.0×10-7). Multiple loci were associated with the presence of neuritic plaques, including APOE (P=1.8×10-30), GALNT7 (P=6.0×10-9), ABCG1 (P=8.0×10-9), and a region near LMX1B (P=4.3×10-8). Additional loci were found to be associated with several ordinal neuropathology traits including LB, AA, MTS, VBI, and AD Braak staging (data to be presented). Conclusion: These results confirm several known AD risk loci and implicate novel loci in the etiology of LOAD neuropathology features, neuritic plaque in particular. Additional analyses including gene-based and pathway-based tests are being performed and will be reported, including thorough investigations of the known LOAD candidate genes.
1314T

Canine Multiple System Degeneration is Associated with Distinct SERAC1 Mutations in Two Different Dog Breeds. J. Guo1, R. Zeng1, G.S. Johnson1, T. Mthianga-Mutangadura1, E. Moravač1, T. Kozic2, J.F. Taylor2, D.P. O’Brien2, R.D. Schnabel2. 1) Department of Veterinary Pathobiology, University of Missouri College of Veterinary Medicine, Columbia, MO; 2) Hayward Genetics Center, Tulane University, LA; 3) Division of Animal Sciences, University of Missouri College of Agriculture, Food and Natural Resources, Columbia, MO, USA; 4) Department of Veterinary Medicine and Surgery, University of Missouri College of Veterinary Medicine, Columbia, MO, USA.

Canine multiple system degeneration (CMSG) is an autosomal recessive neurodegenerative disease. The initial cerebellar ataxia appears at 3 to 6 months and progresses to severe parkinsonism requiring euthanasia during the second year of life. Marked degeneration of the cerebellum, caudate nucleus, thalamus, and substantia nigra are apparent at necropsy. CMSG has been recognized in two breeds: Kerry Blue Terrier (KBT) and Chinese Crested (CC). Previously, we mated CMSG obligate carriers from the two breeds and produced two normal and two CMSG-affected siblings, indicating that in both breeds CMSG is caused by the same mutation or by allelic mutations. Linkage mapping with samples from both breeds restricted the CMSG locus to a 5.2 Mb segment of CFA1 which contained 83 genes. We used illumina sequencing technology with paired-end libraries to generate a whole genome sequence (WGS) with a 24-fold average coverage for a CMSG-affected KBT and a WGS with a 21-fold average coverage for a CMSG-affected CC. The reads from each WGS were separately aligned to the canine reference genome sequence (build 3.1) with NextGene software. The detected sequence variants were prioritized according to predicted functionality and were then functionally assessed in 54 other WGSs generated for dogs not affected with CMSG. After filtering, two homozygous sequence variants from the KBT alignment and four homozygous sequence variants from the CC alignment were in genes located within the previously mapped CFA1 region. Among these were a SERAC1 nonsense mutation in exon 15 in the KBT and a 4 bp deletion in a SERAC1 splice donor site sequence in the CC. RT-PCR amplification of RNA from KBTs and all 20 CMSD-affected CCs were homozygous for their respective CMSD-affected sequence variants. After filtering, all 6 CMSD-affected dogs were homozygous for these mutations, whereas, none of the 287 genotyped CMSD-free dogs were homozygous for these alleles. Mutations in human SERAC1 have been associated with a range of serious neurodegenerative disorders such as MEGDEL syndrome. MEGDEL syndrome patients exhibit dystonia, deafness, Leigh-like syndrome, impaired oxidative phosphorylation, and 3-methylglutaconic aciduria. The pathologic pathways underlying the neurodegeneration in this disease are poorly understood. Dogs with nullifying mutations in canine SERAC1 may prove useful as models to investigate these pathways.

1316W

Serotonin Transporter Methylation and Response to Cognitive Behaviour Therapy in Child Anxiety Disorders. S. Roberts1, K.J. Lester1, C.C.Y Wing1, J.C. Hudson2, G. Creswell3, J. Mill1, T.C. Eley1. 1) King’s College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, UK; 2) Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, NSW, Australia; 3) Winnicott Research Unit, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK; 4) University of Exeter Medical School, Exeter University, St Luke’s Campus, Exeter, UK.

Anxiety disorders are the most common psychiatric disorders in childhood and are associated with a range of social and educational impairments. Cognitive Behaviour Therapy (CBT) is effective for the majority of childhood anxiety cases, although up to 40% of children retain significant impairments after CBT. Recent research in the newly emerging field of ‘therapy genomics’ indicates that individual differences in treatment response may have a genetic basis. Of particular interest is the Serotonin Transporter gene (SLC6A4), which has been associated with disorders such as anxiety and depression. Furthermore, the short (S) allele of the serotonin promoter polymorphism (5HTTLPR) has been associated with poorer outcomes in high-stress environments but better outcomes in positive environments. A study from our group demonstrated that anxious children with the SS genotype showed an increased response to CBT, suggesting an interaction between genes and environmental influences. Epigenetic mechanisms that influence transcriptional regulation (such as DNA methylation) have been shown to be susceptible to the environment, making them plausible candidates for the biological embedding of experience. We tested whether DNA methylation of CpG sites upstream of the SERT promoter were associated with response to CBT in a subsample of 168 clinically anxious children. Buccal swabs DNA was collected at pre- and post-treatment, and SERT methylation determined using the Sequenom EpiTyper. Treatment response was defined as the presence or absence of all anxiety disorder diagnoses following CBT. We detected a significant 3-way interaction between clinical response, percentage methylation at pre- and post-treatment and CpG site (F(5,77)= 2.39, p = 0.046). A significant difference between treatment responders and non-responders was found in percentage methylation change spanning across all sites at follow-up. All 6 CMSG-affected KBTs and all 20 CMSD-affected CCs were homozygous for their respective mutations, whereas, none of the 287 genotyped CMSD-free dogs were homozygous for these alleles. Mutations in human SERAC1 have been associated with a range of serious neurodegenerative disorders such as MEGDEL syndrome. MEGDEL syndrome patients exhibit dystonia, deafness, Leigh-like syndrome, impaired oxidative phosphorylation, and 3-methylglutaconic aciduria. The pathologic pathways underlying the neurodegeneration in this disease are poorly understood. Dogs with nullifying mutations in canine SERAC1 may prove useful as models to investigate these pathways.

1315F

Does gene function predict cognition and behaviour in X-linked intellectual disability? K. Baker1, G. Scent1, D.E. Astle2, P.C. Fletcher4, F.L. Raymond4. 1) Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom; 2) Department of Experimental Psychology, University of Oxford, United Kingdom; 3) MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom; 4) Behavioural and Clinical Neurosciences Institute, University of Cambridge, United Kingdom.

Genetic diagnosis is rapidly becoming a reality for many more individuals and families affected by Intellectual Disability (ID). Post-diagnostic challenges include establishment of a prognostic evidence base for ultra-rare disorders and elucidation of neurodevelopmental mechanisms. To this end, we have collected standardised medical, behavioural and cognitive data on participants with ID in the GOLD (Genetics of Learning Disability) cohort, who carry pathogenic variants in X-linked genes implicated in diverse neuronal functions including receptor dynamics, vesicle trafficking, ubiquitination and palmitoylation. Our aim is to determine which aspects of phenotype may be predicted by gene function. Severity of global impairment (assessed via Vineland Adaptive Behaviour Scales) varies widely within groups. However, beyond global impairment, function-associated characteristics start to emerge. For example, mutations in membrane-associated guanylate kinase genes (e.g. DKG3, PAK3) are associated with poor cognitive tasks (assessed via WASI-II), and high rates of behavioural problems particularly affecting attention and anxiety (assessed via the Developmental Behaviour Checklist). Mutations in pre-synaptic and post-synaptic genes are associated with deficits in accuracy and speed-of-responding on computerised assessments of visual and auditory attention. Lastly, we have identified a previously-unreported association between ZDHHC9 and childhood-onset nocturnal seizures, highlighting the importance of post-diagnostic medical phenotyping. The specificity of these observations will be tested via investigation of larger post-diagnostic samples and additional functional groups.
1317T
Using transcriptomic data to understand the processes involved in antidepressant treatment response. K. Hodgson1, K.E. Tansey1, G. Copola2, G. Breen1, R. Uher1,2, P. McGuffin1, 1) SGDP, Institute of Psychiatry, Kings College London, London, United Kingdom; 2) Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, USA; 3) Department of Psychiatry, Dalhousie University, Halifax, NS, Canada.

Research indicates that antidepressant response is a complex trait, involving interactions between both genetic and environmental factors (Tansey et al 2012). However the mechanisms by which antidepressants have their effects is unclear, and clinicians remain unable to identify who will respond best to which drug. Gene expression levels may capture the molecular changes that occur with antidepressant treatment, and can be influenced by both genetic and environmental factors. By examining the transcriptome and its alteration during treatment, this study aims to understand the biological mechanisms of antidepressant action, and identify if any of these processes are linked to individual variability in treatment response. Blood samples were taken from patients recruited into the Genome-Based Therapeutic Drugs for Depression (GENDEP) study both prior to treatment (week 0) and after receiving an antidepressant for eight weeks. RNA was extracted from whole blood, and processed on the Illumina HumanHT-12 v4 Expression BeadChip. After quality control measures, 123 patients had transcriptomic data available from both time points. Gene probes which varied significantly in expression levels between week 0 and week 8 were identified, and gene probes with expression levels significantly associated with antidepressant response (as measured by percentage change in the Montgomery-Asberg Depression Rating Scale) were identified. Network-level analyses of gene expression changes were undertaken using weighted-gene coexpression network analysis (WGCNA) to examine networks of gene coexpression, for analysis of alterations in gene-connectivity during treatment and links to treatment response. No significant associations were observed when performing the gene probe level analysis, using a threshold of FDR q<0.1. Network level analysis undertaken using WGCNA identified modules of coexpressed genes shared between week 0 and week 8 samples. However, the changes in the gene probe levels within each module (module eigengenes) from week 0 and week 8 did not show significantly over the period of drug treatment. Whilst previous research looking into gene expression changes associated with antidepressant treatment has focused on candidate genes. However, our results indicate that any gene expression changes occurring as a result of antidepressant treatment are not of a large enough magnitude to be detected at transcriptome-wide significance at either the probe or network level.

1318F
MicroRNA-137 is Associated with Epigenetic Variation at the HCG9 Gene in the Amygdala. A.P.S. On1, K.R. van Eijk2, M.F. Stokman2, E. Stronman2, The Netherlands Brain Bank4, M.P.M. Boks5, S. de Jong5, R.A. Ophoff1, 6, 1) Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, CA; 2) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 3) Department of Clinical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 4) The Netherlands Institute of Neuroscience, Amsterdam, The Netherlands; 5) Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, Utrecht, The Netherlands; 6) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.

Noncoding RNAs are widely abundant in cells and fulfill critical roles as transcriptional and post-transcriptional regulators. Noncoding RNAs have been implicated in the regulation of the epigenetic mechanisms. MicroRNA-137 (miR-137) is a small non-coding RNA important for neurodevelopment. A large genome-wide association study recently suggested miR-137 mediated dysregulation as an etiologic mechanism in schizophrenia. This study aims to investigate the role of miR-137 function in epigenetic variation across regions of the human brain.

We investigated 249 postmortem brain samples originating from 44 nondemented control, 6 schizophrenia and 11 bipolar disorder individuals. Data was collected on whole genome DNA methylation and microRNA-137 expression. Linear models were used to investigate the relationship between microRNA expression and DNA methylation levels.

MicroRNA-137 expression showed to be regional specific with highest expression in cortical regions, limbic system and basal ganglia and lowest expression in the cerebellum. In addition, miR-137 expression showed a strong correlation with methylation levels at a CpG site of the HLA complex group 9 gene (HCG9) in the amygdala. This correlation was significant (b = 6.98, t = 9.52, p=0.0066) and showed to be independent of disease status. This study demonstrates a relationship between the schizophrenia associated microRNA miR-137 and DNA methylation at HCG9 in the amygdala. This is a brain region involved in processing emotions and memories and reproducibly implicated in schizophrenia pathology. Interestingly, the HCG9 gene is located within the major histocompatibility complex (MHC) class 1 region, which is also associated to schizophrenia. Future work involves investigating the effect of the miR-137 schizophrenia risk variant and miR-137 target genes on miR-137 expression levels and epigenetic variation at the HCG9 gene. Taken together, this study reveals microRNA-137 expression across the human brain and provides key insights on cross talk between this small RNA molecule and DNA methylation. These findings contribute to understanding etiologic mechanisms in schizophrenia pathology.
1319W
The genetic architecture of white matter integrity: Insights from quantitative genetics and genome-wide associations. E. Sprooten1, E.E.M. Knowles1, D.R. McKay2, M.A. Carless3, M.A.A. de Almeida4, A. Winkler5, T. Dyer5, J.E. Curran2, H. Göring2, R. Olveira2, P. Kochunov6, P. Fox7, L. Almasy7, R. Duggirala1, J. Kent4, J. Blangero8, D.C. Glahn1. 1) Department of Psychiatry, Yale University, New Haven, CT; 2) Department of Genetics, Texas Biomedical Research Institute, University of Texas Health Science Center; 3) FMRIB, Oxford University; 4) Maryland Psychiatric Research Center, University of Maryland; 5) Research Imaging Institute, University of Texas Health Science Center San Antonio.

Background: White matter integrity as measured by diffusion tensor imaging (DTI) is heritable and implicated in several complex phenotypes including major depressive disorder, bipolar disorder, schizophrenia and dementia. We investigated the genetic architecture of white matter and its homogeneity across regions using quantitative genetics and genome-wide association (GWAS) of global and regional fractional anisotropy (FA).

Methods: Participants included 776 Mexican-American individuals from extended pedigrees. DTI data were acquired on a Siemens 3T Trio scanner (b=0 and b=700 s/mm²) along 35 non-collinear directions and a spatial resolution of 1.7x1.7x3mm. TBSi was applied to create white matter skeletons representing the centers of white matter within each subject. Heritability estimates of mean FA within this skeleton and within 48 regions of interest, and genetic correlations between them, were calculated using SOLAR. Genome-wide association was conducted utilizing Illumina microarrays under an additive genetic model. Covariates included sex, age, their interactions, and the first 4 principal components to account for population stratification.

Results: FA was significantlyheritable (h²=0.52, p=1.09*10⁻¹⁰). Genetic correlations between global FA and each of the tracts were between .28 and .94. A number of genome-wide significant SNPs were found for global FA. The strongest association was on 17q24.1 with the intergenic SNP rs10853057 (p=3.18*10⁻¹⁰) close to the gene GNA13, and the second strongest SNP rs12249377 (p=3.26*10⁻¹⁰) was in the serotonin receptor HTR7 gene on 10q23.1. An additional three significant hits were on chromosomes 12 (p=7.97*10⁻¹⁰), 16 (p=1.94*10⁻⁹) and 1 (p=2.00*10⁻⁹). The betas of the GWAS of the global FA measure explained on average 44% (15%-70%) of the rankings of the betas of sub-regions of interest.

Conclusions: FA is a heritable trait and a powerful quantitative phenotype to examine in genome-wide association studies. Genetic correlations and comparisons with other cognitive measures for a substantial portion of genetic influence on white matter structure is uniform across the brain. Our results encourage further research into the biological roles of these genome-wide significant SNPs in health and disease.

1321F
A Meta-analysis including 9,038 individuals confirms an interaction effect between depression and FTO genotype on BMI. M. Rivera1,2, T. Corre1,3, C. Wolf2, D. Czamara2, S. Kloiber2, B. Muller2, M. Preissig3, G. Breen1, C. Craig1, A. Farmer1, C. Lewis1, P. McGuffin1. 1) MRC ŚGD Centre, Institute of Psychiatry, King’s College London, London, United Kingdom; 2) CIBERSAM, University of Granada, Spain; 3) Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; and 4) Swiss Institute of Bioinformatics, Lausanne, Switzerland. 5) Max-Planck-Institute of Psychiatry, Munich, Germany; 6) Department of Psychiatry, CHUV, Lausanne, Switzerland.

Background: Depression and obesity are leading causes of disease burden and disability and major public health concerns worldwide. Both conditions are highly prevalent and are major risk factors for chronic (physical) diseases. The association between depression and obesity has repeatedly been reported in many studies. The role of the FTO gene in BMI and obesity has been confirmed in many independent studies. Recently, we have reported for the first time that depression amplifies the effect of FTO gene variation on BMI. The aim of the present study is to replicate these findings by investigating the interaction FTO rs9939609 polymorphism in a meta-analysis including four independent studies consisting of depression cases and controls.

Methods: The sample consists of 5,134 depression cases and 3,904 controls from different studies: Radiant (3,251), PsyCoLaus (2,994), GSK (1,577) and MARS (1,116). As common inclusion criteria were available information on depression phenotypes, BMI and genotype data for rs9939609 FTO polymorphism. The distribution of BMI was positively skewed in all studies. We therefore transformed the data to Log₁₀(BMI) to achieve a closer approximation to normal distribution. In each individual study, linear regression models for quantitative traits assuming an additive genetic model were performed to test for the interaction between rs9939609 polymorphism and depression for an effect on Log₁₀BMI. All individuals were of white European ancestry. A classical approach meta-analysis with interaction terms was used and standard errors were obtained using the statistical package METAL. Results: The results supported a significant interaction between FTO rs9939609 genotype and depression in relationship to Log₁₀BMI (β=0.092, SE=0.033, p=0.0058). There was significant heterogeneity among studies. The meta-analysis interaction results show that in cases with depression there is an increased of 0.092 units of BMI for each FTO C allele.

Discussion: This is to date the first meta-analysis investigating the relationship between FTO and depression. The results confirm that a history of depression increases the effect of FTO gene on BMI. This finding could have implications for predicting which patients with depression are at risk of high-BMI related disorders and potentially highlights how to improve prevention, management and treatment programs.

1320T
Association of KIBRA With Episodic and Working Memory: A Meta-Analysis. T. Dyer1, M. Almasy2, R.Duggirala1, J.E. Curran1, S. Kloiber2, B. Muller2, M. Preissig3, G. Breen1, A. Farmer1, C. Lewis1, P. McGuffin1, C. Craig1, A. Farmer1, C. Lewis1, P. McGuffin1. 1) MRC ŚGD Centre, Institute of Psychiatry, King’s College London, London, United Kingdom; 2) CIBERSAM, University of Granada, Spain; 3) Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; and 4) Swiss Institute of Bioinformatics, Lausanne, Switzerland. 5) Max-Planck-Institute of Psychiatry, Munich, Germany; 6) Department of Psychiatry, CHUV, Lausanne, Switzerland.

WWC1 was first implicated in human cognition through a genome wide association study in 2006 that reported an association of the intronic single nucleotide polymorphism (SNP) rs17070145 with episodic memory performance. WWC1 encodes the protein KIBRA, which is almost ubiquitously expressed. Together with its binding partners, KIBRA is assumed to play a role in synaptic plasticity. T-allele carriers of SNP rs17070145 have been reported to outperform individuals that are homozygous for the C-allele in many tasks in samples of primarily Caucasian background.
1322W
The genetic basis of solitude: lower heterozygosity in people with strict definition autism spectrum disorders and their parents. S. Huang, X. Lu, J. Liang, D. Yuan. State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China.

Autism Spectrum Disorders (ASDs) are >4 times more common in males and parents of people with ASDs often show mild forms of autistic-like characteristics or broad phenotypes. Strict (STR) and spectrum (SPC) definition ASDs differ mainly in social deficits with STR only slightly lower in IQ. Our recent work reveals a link between nucleotide diversity and complex traits/diseases consistent with stabilizing selection on SNP amounts and a priori truth on entropy and order. Heterozygosity (Het), a measure of random diversity, may be under stabilizing selection. From a priori reasoning, solitude should be linked with more homozygous (Hom) genotypes since social deficits and tendency to marry people like oneself favor less admixture. Parents of ASDs are more populated in science and engineering fields demanding both IQ and creativity. While ASDs may be under negative selection, solitude is critical to creativity. An aesthetic hallmark of great creative works is unity in harmony of opposites in a pure and high contrast fashion. A genome ranked high in aesthetics with more Hom parts should make more creative people with all around balanced traits. Remarkably, previous work indeed shows less social but more novelty seeking behavior in inbred versus hybrid mice. We analyzed SNP dataset from dbGAP of 1154 ASD trios of European ancestry and correlated each individual's number of Het SNPs with STR vs SPC status. We found a moderate but significant correlation of low Het to STR cases (Spearman r = 0.086, P < 0.01) and their fathers (r = 0.08, P < 0.01). The average Het of STR cases is lower than SPC (P < 0.05, t test). STR sons have lower Het than their Hom minor allele content or amount (MAC) than their fathers (P = 0.02, t test), and the same holds for SPC sons and fathers (P = 0.006). (MAF was determined using parents.) Similar results were found for 148 trios of non-European ancestry. Finally, we analyzed relevant control cohorts from dbGAP including the Framingham study. Low Het associates significantly with living alone in never married, single and older age, and higher education level and income. These results suggest that function and solitude in STR fathers may be at precarious near imbalance under more Hom DNAs. A moderate decrease in their sons in Hom MAC coupled with a slight increase in Het may tilt the balance towards extreme functions at either end of the spectrum.

1323T
Shared genetic vulnerability for Attention Deficit Hyperactivity Disorder, Substance Use and Gambling in Australian Adolescents. P.A. Lind1, D.A. Hay2, T.J. Cicero3, N.G. Martin4, S.E. Medland1. 1) Quantitative Genetics, Queensland Institute of Medical Research, Brisbane, Australia; 2) School of Medicine, Department of Psychiatry, Washington University, USA; 3) Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia.

Previous studies have shown that the prevalence of substance use (including legal and illicit drugs) and gambling in children with Attention Deficit Hyperactivity Disorder (ADHD) is higher than in the general population. The aim of this study was to explore whether cigarette smoking, alcohol drinking and gambling in adolescents with ADHD was due to shared genetic risk factors. We analyzed self-report substance use and gambling data in twin participants from the Genetics of Inattention in Australia study (N = 500 individuals) who completed a structured psycho-social interview. Data included age of first use, number of DSM-IV symptoms for substance dependence or pathological gambling, the maximum number of cigarettes smoked in 24 hr (MAXCIG), the maximum number of alcoholic drinks consumed in 24 hr (MAXDR), and the number of different types of gambling used (PGVAR). Using quantitative genetic modeling, we found that genetic risk factors for ADHD also increased DSM-IV dependence symptoms (genetic correlations, r = .03–.19), the number of cigarettes consumed (r = .45) and types of gambling tried (r = .07), and reduced the age of onset for gambling, smoking and using alcohol (r ranges from −.07 to −.48). However, a negative genetic correlation was observed with alcohol consumption (r = −.29).

These results suggest that the variants influencing ADHD behaviors influence substance use and gambling in later life and provide support for the development of an ADHD-/substance use cross disorder consortium.

1324F
Genome-wide association study for Cognitive Decline. L.B. Chibnik1, 2, L. Yu2, T. Raj3, T. Xu4, N. Patsopoulos5, 6, B.T. Keenan7, R. Sherva8, S.E. Largent2, D. Blacker2, R.S. Wilson1, E.M. Reiman9, M. Huentelman8, 9, R.C. Green10, 11, L.A. Farrer8, P. Crane11, R. Mayeux12, 13, R. Lipton14, G.D. Schellenberg15, D.A. Evans16, P.L. De Jager3, 4, D.A. Bennett5, Alzheimer's Disease Genetics Consortium. 1) Department of Neurology, Harvard Med Sch/Brigham & Women's Hosp, Boston, MA; 2) Program for Functional and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL; 4) Boston University, School of Medicine, Boston, MA; 5) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 6) Neurogenomics Division, Translational Genetics Research Institute and Arizona Alzheimer's Consortium, Phoenix, AZ; 7) Banner Alzheimer's Institute and Department of Psychiatry, University of Arizona, Tucson, AZ; 8) Department of Genetics, Department of Medicine, Brigham & Women's Hospital, Boston, MA; 9) Partners Healthcare Center for Personalized Genetic Medicine, Boston, MA; 10) Department of Neurology, Columbia University, New York, NY; 11) GerHed Science, Columbia University, New York, NY; 12) Departments of Neurology, Epidemiology and Population Health and the Montefiore Headache Center, Albert Einstein College of Medicine, Bronx, NY; 13) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 14) Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, IL.

Background: Cognitive decline, especially decline in episodic memory, is the clinical hallmark of Alzheimer’s disease (AD) and known to start long before the onset of clinically diagnosed AD dementia. Thus, it may serve as a useful quantitative phenotype for GWAS and other ‘omics’ analyses. Methods: We performed a genome-wide association study (GWAS) on a large prospective and community-based cohort studies (ROS, MAP, CHAP, EAS) who were non-demented at study entry and have at least two repeated measures of cognition. Within each cohort, individual cognitive tests were combined to form aggregate measures of global cognition and episodic memory. Genotype data from each cohort was quality controlled and imputed using 1000Genomes reference panel. We used linear mixed effects models to characterize individual paths of change in cognition, modeling age, sex and education as fixed effects and intercept and slope as random effects. We extracted individual residual cognitive slope estimates from the models and fit a linear regression model for each single nucleotide polymorphism (SNP), adjusting for population stratification. Results: Meta-analyses across cohorts. Results were meta-analyzed across four distinct analyses. The average age at enrollment in each study ranged from 72 to 81, follow-up ranged from 2–18 years and incidence of clinical diagnosis of AD ranged from 19% and 34%. The most significant loci for decline in both global cognition and episodic memory was APOE (p=3.1×10−34 and p=2.1×10−24, respectively). Although no other non-APOE locus reached genome-wide significance, we found suggestive results at two other regions. First, near the EDAR gene, previously found suggestive in a large AD case/control GWAS, with the top SNP showing significant association with cognitive decline (rs13506954, p=8.9×10−8 for global cognition and p=1.0×10−4 for episodic memory). Second, on the RAB3GAP2 gene with the top locus (non-coding intron) reaching a p=1.3×10−7 for global cognition and p=1.3×10−6 for episodic memory. We are currently analyzing data from two other cohort studies for validation. Conclusion: One previously un-described SNP and one possible AD associated SNP are observed to be associated with cognitive decline in a meta-analysis of community-based elders followed longitudinally. Our findings to date suggest that cognitive decline may be a useful phenotype for identifying genomic variation associated with AD or other factors influencing cognitive decline.

1325N

Diabetic retinopathy (DR) is a leading cause of vision loss in the United States (US) and is the most common cause of blindness worldwide. TGFBI, the gene encoding the extracellular matrix protein laminin-5, is expressed in the retina and has been implicated in the development of retinal neovascularization in DR. We hypothesized that TGFBI is involved in the pathogenesis of retinal neovascularization in DR. Therefore, we designed the following experiments: 1) We performed a genome-wide association study (GWAS) for retinal neovascularization in DR using a publicly available dataset of 18,000 individuals with DR. We identified a novel SNP (rs1123505) in the TGFBI gene associated with retinal neovascularization in DR (p=4.4×10−6). 2) We performed an in vitro study using human retinal pigment epithelial cells (hRPE) and human retinal vessels (hRV) to investigate the role of TGFBI in retinal neovascularization. We found that TGFBI expression was increased in hRPE and hRV from patients with DR compared to controls (p<0.05). Furthermore, we found that TGFBI increased the number of angiogenic sprouts from hRPE and hRV in a dose-dependent manner. 3) We performed a computational analysis using machine learning algorithms to predict the expression of TGFBI in retinal neovascularization. We found that TGFBI expression was predictive of retinal neovascularization in DR with an area under the curve of 0.79. In conclusion, our findings suggest that TGFBI plays a role in the development of retinal neovascularization in DR. Further studies are needed to determine the biological mechanisms underlying this association.
1325W Combining RNA-sequencing and genotyping data to identify genes and pathways associated with major depression. S. Mostafavi1, A. Battle1, X. Zhu1, J. Potash2, M. Weissman3, J. Shin4, K. Beckham4, C. Haudenschild5, C. McCormick2, R. Me2, M. Gameroff2, H. Glindes3, P. Adams2, F. Goes2, F. Mondimore6, D. MacKinnon7, L. Notes10, B. Schweizer2, D. Furman1, S. Montgomery4, A. Urban5, D. Kollier1, D. Levinson1. 1) Stanford University, Stanford; 2) University of Iowa, Iowa City; 3) Columbia University, NY; 4) National Cancer Institute, Bethesda; 5) University of Minnesota, Minneapolis; 6) Personalis, Menlo Park; 7) Illumina, La Jolla; 8) Centrofli Biosciences, Palo Alto; 9) Johns Hopkins University, Baltimore; 10) American University, Washington, DC.

Altered gene expression levels in disease can reflect the effect of sequence variation, environmental factors and their interaction with genetics, as well as the effects of the disease processes itself. To gain insights into biological mechanisms underlying major depressive disorder (MDD), we identified genes and pathways associated with major depression (MDD) by combining RNA-seq and genotyping data. By analyzing expression and genotype data from subjects with MDD and healthy controls, we were able to identify genes and pathways that are associated with MDD. This approach allows us to identify genetic factors that contribute to the development of MDD and potentially help in the diagnosis and treatment of the disease.

1327F Adult onset painful polyneuropathy caused by a dominant NAGLU mutation. M. Tetreault1,2, MJ. Dicaire3, P. Allard1, K. Gehring4, D. Leblanc1, N. Leclerc3, J. Mathieu3, B. Brais1. 1) Neurogenetics Laboratory, Neurology and Neurosurgery, Montreal Neurological Institute, Montreal, PQ, Canada; 2) Centre of Excellence in Neurosciences of Université de Montréal, CHUM Research Center, Montréal, Québec, Canada; 3) Laboratoire de génétique médicale, CHU-Ste-Justine, Montréal, Québec, Canada; 4) Department of Biochemistry, McGill University, Montréal, Québec, Canada; 5) Cliniques de maladies neuromusculaires, Jonquière, Québec, Canada.

Late-onset painful sensory neuropathies are usually considered to be acquired and most commonly associated with common diseases such as diabetes. Adult presentations of known hereditary forms such as Mucopolysaccharidosis Fabry’s are accompanied by other organ involvement. We recruited a large French-Canadian family (45 individuals; 21 affected cases) with a dominantly inherited late onset painful sensory neuropathy to uncover the underlying genetic mutation. The main clinical features of this dominant painful sensory polyneuropathy are the appearance of excessive cramps as early as 20 years old, constant painful paresthesias in the feet and later the hands appearing on average around age 55 (30–65) that interfere with sleep and progresses into a mild sensory ataxia. Electrophysiological studies are normal until late in the course of the disease where it documents a sensory polyneuropathy. Four affected individuals were enrolled for exome sequencing. Analysis of rare variants shared by all affected cases led to a list of four candidate variants. Segregation analysis in all recruited individuals has shown that only the p.Ile403Thr variant in the α-N-acetyl-glucosaminidase (NAGLU) gene is segregating with the disease. Recessive NAGLU mutations cause the severe childhood lysosomal disease Mucopolysaccharidosis IIIB (MPS-IIIB). Family members carrying the mutation showed a significant decrease of the enzymatic function. The late onset and variable severity of the symptoms may have precluded the description of such symptoms in parents MPS IIIB cases since they usually are lost to follow-up. The identification of a dominant phenotype associated with a NAGLU mutation supports that some carriers of lysosomal enzyme mutations may develop later in life milder phenotypes.


Autism is a complex neurodevelopmental disorder behaviorally characterized by impairments in social interactions, communication and repetitive and restricted patterns of behavior. Autism has a strong genetic component. Currently more than 100 genetic causes of autism are recognized accounting for approximately 20% of cases and indicating high levels of genetic heterogeneity. Recently additional genes with de-novo mutations with large effects were identified in simplex families, with each gene accounting for a very small fraction of cases. In similar manner to genetics of simplex autism it is likely that for a subset of familial cases rare highly penetrant genetic variants are causal as well. Identification of genes responsible for familial autism is important as genetic architecture of familial autism might have important epidemiological implications. Under assumption that causal mutations are private or very rare, exome sequencing is a powerful tool for identification of highly penetrant variants in families with multiple affected cases. We performed whole exome sequencing in 27 NIMH and University of Washington autism families. We have selected families that include affected cousins and frequently additional affected siblings to minimize variant sharing between affected family members. The exome sequencing using NimbleGen SeqCap v2.0 and Illumina HiSeq2000 was performed at University of Washington Northwest Genomic Center. Sequences were aligned to the reference genome with BWA and variants were called with GATK. Functionally annotations of the variants were performed using ANNOVAR. As candidate variants we consider variants that are private, (not present in dbSNP, 1000 genomes or ESP6500) and functional (coding, missense, stop-gain, stop-loss, splice). The list of annotations and frequency filtering resulted in 424 genes with single de-novo mutations from recent exome sequencing studies. We have found 7 genes that are present in de-novo mutations in 22±12 shared private functional variants. In aggregate our sample has generated 198 private candidate variants. We have compared our candidate gene list with list of 424 genes with single de-novo mutations from recent exome sequencing studies. We have found that only 1% of the genes present in de-novo mutations in single de novo mutations is shared between our two datasets. As a next step, to confirm association with autism for the genes with mutations in simplex and familial cases, we are using molecular inversion probes and next generation sequencing to perform a large based case-control study of 1500 familial autism cases and 500 unscreened controls.

1326T The PD Brain Map Project: Mapping the Transcriptional Architecture of Dopamine Neurons in Human Brain. C.R. Scherzer1,2, C. Vanderburg3, T.B. Funder2, D. Nickerson3, Z. Liu4, B. Zheng4, Z. Liao1. 1) The Neurogenomics Laboratory, Harvard Medical School and Brigham & Women’s Hospital, Cambridge, MA; 2) Department of Neurology, Massachusetts General Hospital, Boston, MA; 3) Civin Laboratory for Human Genetics and Developmental Biology, Banner Sun Health Research Institute, Sun City, AZ; 4) Department of Neurology, Arizona Mayo Clinic, Scottsdale, AZ; 5) The Broad Institute of MIT and Harvard, Cambridge, MA; 6) Department of Neurology, Brigham & Women’s Hospital, Boston, MA.

While the number of human genes has shrunk to an estimated ~22,000, a hidden universe of an ever-increasing number of tens of thousands of non-coding RNAs and splice variants is revolutionizing our thinking about the complexity of the human brain. We hypothesize that in complex genetic neurodegenerative diseases such as Parkinson’s (PD) combinatorial effects of genetic and environmental risks disrupt the ordered flow of genetic information into vulnerable brain cells through cell-type-specific modulation of transcript abundance, transcription site use, and sequence. The PD Brain Map Project’s goal is to chart the flow of information from the entire genome to the cell-type specific mechanisms in dopamine neuron. Genetic variation between more than 100 individuals is examined for association with differences in transcribed elements — both protein-coding and non-coding — to identify regions of the genome that influence whether, how, and how much a transcript is expressed in this specific cell type in situ in human brains. Transcriptomes of control and diseased brains are probed using laser-capture microdissection, massively parallel sequencing of near ultra-low amounts of RNA, and expression Quantitative Trait Locus analysis. The PD Brain Map Map will include single cell genetic and epigenetic elements in a prototype cell-type in human brains, help to understand inherited susceptibility to PD, and highlight targets for precision therapies. Pilot phase results indicate that tens of thousands of isoforms of protein-coding genes and tens of thousands of non-coding RNAs and splice variants are expressed in dopamine neurons and suggest a more diverse and complex transcriptional architecture than previously imagined. Support: U01 NS082157; WB1XWH-BAAM-12-1; Michael J. Fox Foundation; U24 NS702026, P30 AG19610.

1) Dept of Neurobiological Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands; 2) Dept of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; 3) Institute of Human Genetics, Johannes Gutenberg University Mainz, Mainz, Germany.

The RBFOX1 gene is a muscle, heart and brain specific RNA-binding protein that regulates the alternative splicing of various genes in neuronal development and maintenance. Copy number variations within the RBFOX1 gene (alias FOX1, A2BP1, OMIM *605104) are increasingly realized as causes of human neurodevelopmental disorders including intellectual disability, autism spectrum disorder (ASD), epilepsy, attention deficit hyperactivity disorder (ADHD), and schizophrenia. Here, we report on a 16-year-old male with ASD, mild intellectual disability (IQ 59, tested at age 13 years) and a heterozygous deletionspanning 100.3 kb in the 5’ region of the RBFOX1 gene. From age 2 years onwards, he showed speech delay and behavioural abnormalities including autoagression, head banging and teeth grinding. He responded towards touch with beating, frequently cried himself to sleep and had fits of anxiety, e.g. towards noise. He also had frequent fits of laughter and tantrums without obvious cause. Neurological examinations at 5 and 6 years of age indicated mild muscular hypotonia, reduced gross and fine motor skills, clumsiness and ADHD (agitation, inability to concentrate, distractibility, impulse control disorder, and emotional lability). At age 16 years he read fluently with proper intonation and understanding (school grade 3 level) but could not write, perform simple mathematical calculations or hold eye contact. The deletion on chromosome 16p13.2 (chr16:6,751,824-6,852,175, UCSC hg19) included the promoter region and exon 1 of transcript variant 6 of RBFOX1. A previous report describes a patient with remarkably similar molecular and clinical findings (Mikhal et al., 2011, AJMG-A 155A:2386-2396). Our study further supports the role of RBFOX1 gene mutations and in particular, of heterozygous loss-of-function mutations of RBFOX1 transcript variant 6, in the etiology of ASD, ADHD and intellectual disability.


1) Dept of Neurobiological Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands; 2) Dept of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; 3) Institute of Human Genetics, Johannes Gutenberg University Mainz, Mainz, Germany.

The RBFOX1 gene is a muscle, heart and brain specific RNA-binding protein that regulates the alternative splicing of various genes in neuronal development and maintenance. Copy number variations within the RBFOX1 gene (alias FOX1, A2BP1, OMIM *605104) are increasingly realized as causes of human neurodevelopmental disorders including intellectual disability, autism spectrum disorder (ASD), epilepsy, attention deficit hyperactivity disorder (ADHD), and schizophrenia. Here, we report on a 16-year-old male with ASD, mild intellectual disability (IQ 59, tested at age 13 years) and a heterozygous deletion spanning 100.3 kb in the 5’ region of the RBFOX1 gene. From age 2 years onwards, he showed speech delay and behavioural abnormalities including autoagression, head banging and teeth grinding. He responded towards touch with beating, frequently cried himself to sleep and had fits of anxiety, e.g. towards noise. He also had frequent fits of laughter and tantrums without obvious cause. Neurological examinations at 5 and 6 years of age indicated mild muscular hypotonia, reduced gross and fine motor skills, clumsiness and ADHD (agitation, inability to concentrate, distractibility, impulse control disorder, and emotional lability). At age 16 years he read fluently with proper intonation and understanding (school grade 3 level) but could not write, perform simple mathematical calculations or hold eye contact. The deletion on chromosome 16p13.2 (chr16:6,751,824-6,852,175, UCSC hg19) included the promoter region and exon 1 of transcript variant 6 of RBFOX1. A previous report describes a patient with remarkably similar molecular and clinical findings (Mikhal et al., 2011, AJMG-A 155A:2386-2396). Our study further supports the role of RBFOX1 gene mutations and in particular, of heterozygous loss-of-function mutations of RBFOX1 transcript variant 6, in the etiology of ASD, ADHD and intellectual disability.

The RBFOX1 gene is a muscle, heart and brain specific RNA-binding protein that regulates the alternative splicing of various genes in neuronal development and maintenance. Copy number variations within the RBFOX1 gene (alias FOX1, A2BP1, OMIM *605104) are increasingly realized as causes of human neurodevelopmental disorders including intellectual disability, autism spectrum disorder (ASD), epilepsy, attention deficit hyperactivity disorder (ADHD), and schizophrenia. Here, we report on a 16-year-old male with ASD, mild intellectual disability (IQ 59, tested at age 13 years) and a heterozygous deletion spanning 100.3 kb in the 5’ region of the RBFOX1 gene. From age 2 years onwards, he showed speech delay and behavioural abnormalities including autoagression, head banging and teeth grinding. He responded towards touch with beating, frequently cried himself to sleep and had fits of anxiety, e.g. towards noise. He also had frequent fits of laughter and tantrums without obvious cause. Neurological examinations at 5 and 6 years of age indicated mild muscular hypotonia, reduced gross and fine motor skills, clumsiness and ADHD (agitation, inability to concentrate, distractibility, impulse control disorder, and emotional lability). At age 16 years he read fluently with proper intonation and understanding (school grade 3 level) but could not write, perform simple mathematical calculations or hold eye contact. The deletion on chromosome 16p13.2 (chr16:6,751,824-6,852,175, UCSC hg19) included the promoter region and exon 1 of transcript variant 6 of RBFOX1. A previous report describes a patient with remarkably similar molecular and clinical findings (Mikhal et al., 2011, AJMG-A 155A:2386-2396). Our study further supports the role of RBFOX1 gene mutations and in particular, of heterozygous loss-of-function mutations of RBFOX1 transcript variant 6, in the etiology of ASD, ADHD and intellectual disability.
1333F  Mapping the schizophrenia brain: Genes harboring damaging de novo mutations in schizophrenia map to a highly interconnected network of transcriptional co-expression and protein interaction in fetal prefrontal cortex. S. Gulsunen1, T. Walsh1, A.C. Watts1, M.K. Lee1, A.M. Thornton1, S. Casadei1, C. Rippey1, H. Shahn1, M.-C. King1, J.M. McClellan1, Consortium on the Genetics of Schizophrenia (COGS), PAARTNERS Study Group1, 1Department of Medicine (Medical Genetics), University of Washington, Seattle, WA; 2Department of Psychiatry, University of Washington, Seattle, WA.

Schizophrenia is a chronic neuropsychiatric disorder characterized by both motor and vocal tics. Family studies and twin studies provide strong evidence for a genetic component, genetic causes of the disease remain elusive. Many patients have no family history of mental illness. Genes responsible for schizophrenia can be revealed by de novo mutations in such patients. Participants in our project were 105 quadriplets or trios comprising a proband with schizophrenia, an unaffected sibling if available and their unaffected parents. Families were selected for absence of history of serious mental illness. Exome sequencing of all 399 persons revealed 57 damaging de novo mutations in 47 of 105 schizophrenia probands and 35 such mutations in 25 of 84 unaffected siblings (X2 = 4.45; P = 0.035). The genes harboring de novo damaging mutations yielded networks based on protein-protein interactions (PPI) and transcriptional co-expression in different brain regions at different developmental stages. PolyPhen algorithm analysis from the BrainSpan Atlas of Developing Brain. The network generated by genes harboring de novo damaging mutations in schizophrenia had significantly more physical interactions (P=0.0005) and significantly greater levels of transcriptional co-expression in dorso lateral (P=0.0001) and ventrolateral (P=0.0004) regions of the prefrontal cortex during fetal development, compared to networks derived from genes mutant in unaffected siblings. Overall, 40 genes with de novo damaging mutations in schizophrenia mapped to this highly interconnected network. These genes function in the brain in neuronal migration, synaptic transmission, signaling, transcriptional regulation, and transport. These results support prefrontal cortex as a critical region in the pathogenesis of schizophrenia. The approach also supports the applicability of proteomic and transcriptome analyses in order to map critical genes to shared biological networks for conditions characterized by extreme genetic heterogeneity.

1334W  Targeted re-sequencing of HDC and SLITRK1 in Tourette syndrome using next-generation sequencing. A. Inai1, H. Kuwabara1, Y. Eniguchi1, T. Shimada1, M. Furukawa1, T. Sasaki2, M. Tochigi2, C. Kakukichi2, K. Kasa2, Y. Kano1, 1) Department of Child Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Japan; 2) Department of Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Japan; 3) Department of Health Education, Graduate School of Education, University of Tokyo, Tokyo, Japan.

Introduction; Tourette syndrome (TS) is neurodevelopmental disorder characterized by both motor and vocal tics. Family studies and twin studies provide strong evidence for genetic nature of TS. Although heredity of TS is high, genetic factors remain largely unknown. Association studies revealed that common genetic variants explain a modest fraction of heritable risk. Histidine decarboxylase gene (HDC) and Slit and Trk-like 1 gene (SLITRK1) were implicated as possible susceptibility genes by several association studies. However, the direct contribution of these genes to TS pathogenesis is still unclear.

In the present study, we aimed to investigate whether TS is related to the HDC and SLITRK1 genes using targeted re-sequencing methods. A total of 95 TS cases and 95 controls participated in this study. All the participants provided written informed consent after they were given a complete explanation of the study as required by the ethics committee of the University of Tokyo Hospital. The exon region of HDC and SLITRK1 were amplified on TruSeq Custom Amplion kit following the manufacturer’s protocols. Amplified libraries were sequenced on the Illumina Miseq as pair-end 151-bp reads. Sequence reads were mapped to the reference genome (hg19). Variant calls were filtered to coordinate with at least 10x coverage. Results; There were total of 570 single nucleotide variants (SNVs) called in the 95 TS participants. However, filtering for synonymous substitutions and database SNPs (dbSNP), four novel and non-synonymous variants remained. Three participants had distinct SNVs on HDC, and other three participants had identical SNV on SLITRK1. PolyPhen algorithm predicted that all three SNVs on HDC disrupted the conformation of the protein. On the other hand, the SNV on SLITRK1 was predicted to be benign mutation. Conclusions; Present study suggested that HDC gene could play a role in the pathogenesis of TS, whereas SLITRK1 might play few or no role in TS susceptibility in Japanese population.

1335F  Exome sequences of multiplex, multigenational families reveal schizophrenia risk loci involved in fatty acid oxidation. M.Z. Kos1, J. Peralta1, M.A. Carlesso1, M. Almeida1, R.C. Gur2, M.F. Pogue-Geile3, D. Roa4, V. Nimsoenkar4, R.E. Gur5, L. Almasyk1, 1) Dept Gen, Texas Biomedical Research Institute, San Antonio, TX; 2) Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA; 3) Department of Psychology, University of Pittsburgh, Pittsburgh, PA; 4) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Schizophrenia is a serious mental illness defined by severe psychosis and disruptions in thought and behavior, with a worldwide prevalence of about 1%. The genetic basis of schizophrenia is highly heritable. Although heritability as estimated around 0.80, much of the genetic liability is yet to be explained. In this paper, we search for susceptibility loci in multiplex, multigenational families affected by schizophrenia, by targeting coding regions that are more likely to contain disease-causing variants. The exome sequencing was performed on 134 samples from eight European-American families, including 25 individuals with DSM-IV schizophrenia or schizoaffective disorder, using Illumina TruSeq technology. The exome variant calls were filtered based on stringent GATK probabilistic quality scores (LODs ≥ 4.0), as well as functional relevance as computed by the SIFT and PolyPhen algorithms. In total, 11,878 non-synonymous variants with putative deleterious effects, representing 5,867 genes, were tested for their association with the schizophrenia spectrum disorder. Our association hit is for a common polymorphism (rs1094111; MAF = 0.40) in the gene AMACR, which has been previously implicated in the risk of schizophrenia, a result that remained highly significant after correcting for multiple-testing (corrected P = 0.0043). The enzyme coded by this gene, alpha-methylacyl-CoA racemase, is involved in the metabolic pathways of fatty acid oxidation. Triglyceride and fatty acid deficiency can lead to the accumulation of pristanic acid in neural tissue, which may influence brain function and structure. In addition, the SNP rs10378 in the gene TMEM176A is also significant (corrected P = 0.031). Interestingly, transcriptional expression (TREX) analysis of the hits (P = 0.01; n = 249 genes) revealed significant enrichment of genes involved in "fatty acid oxidation" (empirical P = 0.027), which includes the gene AMACR. Rare, non-synonymous variants were found in two other genes involved in this metabolic pathway: a novel gain-of-function mutation in the gene ACAAT (P = 0.0016; MAF = 0.0088) and rs2894359 in the gene HSD17B4 (P = 0.0065; MAF = 0.025).

In conclusion, our association findings from family-based exome sequence data suggest an important role for genes involved in fatty acid metabolism in the pathogenesis of schizophrenia. In particular, the gene AMACR, providing key insights into possible strategies for preventative and therapeutic treatments.
1337W
A Novel Autosomal Dominant Dystonia and Spastic Paraplegia Caused by a Mutation in ATP5G3, a gene that encodes for subunit c of mitochondrial ATP Synthase. D.E. Neilson1, T. Huang1, X. Wang1, N.D. Leslie1, R.B. Hufnagel1, D.L. Gilbert2. 1) Div. of Human Genetics, Cincinnati Children’s Hospital, Cincinnati, OH; 2) Div. of Neurology, Cincinnati Children’s Hospital, Cincinnati, OH.

OBJECTIVE: To identify the genetic basis for the neurological phenotype in a large, multigenerational family with a novel, autosomal dominant, highly penetrant neurological disease causing childhood onset dystonia and adulthood onset spasticity.

BACKGROUND: We previously characterized the phenotype of this disease which has highly variable onset between the first year and fifth decade. Clinically it presents with progressive, deep brain stimulation-responsive idiopathic generalized dystonia in childhood and/or spastic paraplegia in adulthood. Linkage to chromosome 2q24-2q31 was identified, in a region in which no prior inherited dystonias or spastic paraplegias had been identified.

METHODS: Forty family members were seen and examined of whom 18 were affected. Exome sequencing was performed on two distantly related family members. The resulting analysis was constrained to the genetic locus on chromosome 2q. Confirmatory Sanger-based sequencing was performed on the remainder of the affected and unaffected family members. Functional analyses of isolated mitochondrial complexes, derived from patient fibroblast culture, were performed.

RESULTS: A single, novel heterozygous mutation (p.1606N→K) in ATP5G3 was identified in a highly conserved region of the subunit c subunit of mitochondrial complex V ATP synthase. This mutation segregated with disease in the family. ATP synthase catalyzes ATP synthesis during oxidative phosphorylation. The c subunits of this complex form the trans-membrane proton pore structure. ATP5G1, ATP5G2 and ATP5G3 provide functional redundancy. Mitochondrial assays demonstrate a statistically significant 20% reduction in complex V activity, whereas complexes I through IV displayed normal activity.

CONCLUSIONS: An amino acid substitution in complex c of mitochondrial ATP synthase causes progressive generalized dystonia and spastic paraplegia. Initial studies suggest that the mutant protein interferes with complex V activity, proportional to its relative expression. Tissue specific or stress-induced increases in expression of this mutant ATP5G3 may produce stronger reductions in activity. This may explain the restricted pattern of neuronal dysfunction. In vivo and in vitro studies currently in progress may reveal a mechanism for neurological dysfunction in this and other progressive neurological diseases.

1338T
Whole exome sequencing of patients with Rett-like features negative for MECP2 mutations: H.E. Olson, O. Khwaja, C. LaCoursiere, E. Martin, W.E. Kaufmann, A. Poduri. Neurology, Boston Children’s Hospital, Boston, MA, USA.

Objective: To identify genetic etiologies in cases with Rett syndrome or with Rett-like features when clinical testing for MECP2 mutations or deletions is negative. Methods: A cohort of eleven patients with Rett syndrome-like features, four meeting criteria for the disorder, and negative clinical testing for mutations or deletions in MECP2 were recruited by a Rett syndrome specialist to the Core for Neurological Diseases at Boston Children’s Hospital. We completed a detailed phenotypic analysis and performed whole exome sequencing. Results: Using 2010 diagnostic criteria, three patients had classical Rett syndrome and mutations in MECP2 (two frameshift deletions and one pathogenic missense mutation). One patient met criteria for atypical Rett syndrome, with neonatal onset epilepsy including focal seizures and epileptic spasms, and had a frameshift deletion in STXBP1. The remaining patients had Rett-like features but did not meet criteria for Rett syndrome, most often due to lack of regression. Patients with Rett-like features without epilepsy had a missense mutation in FOXL1, and consistent MRI findings. One had a deletion in MECP2. For the remaining five, candidate genes were identified including known epilepsy genes. Conclusions: Whole exome sequencing is high yield for patients with Rett syndrome or Rett-like features negative for mutations in MECP2, though targeted gene testing may also provide a diagnosis. Genes associated with atypical Rett syndrome, epilepsy, or intellectual disability should be considered in cases not meeting criteria for Rett syndrome or when MECP2 testing is negative. Clinical criteria supportive of classical Rett syndrome correlated well with MECP2 mutations.

1339F
Detecting novel mutations of Alzheimer’s disease gene using semicon- ductor sequencing. Y. Ryoichi1,2, M. Ryosuke1,2, M. Hiroyuki1, Y. Yushin3, K. Masahito4,5, K. Takashi1, M. Hirofumi1, M. Noriyoshi2, K. Hidemi2, K. Hidemi1. 1) Epidemiology, Research Institute for Radiation Biology and Medicine Hiroshima University, Hiroshimashi, Hiroshimaken, Japan; 2) Department of Periodontal Medicine, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan; 3) Department of Neurology, Tokushima University Hospital, Tokushima, Japan; 4) Department of Anatomical Pathology, Hiroshima University Hospital, Hiroshima, Japan.

Alzheimer’s disease (AD) is a progressive neurodegenerative disease, characterized by an impaired ability to remember, poor judgment, impaired visuospatial abilities, language functions or changes in personality. Neuronal loss, senile plaque, and neurofibrillary tangle are pathological hallmarks of AD. To date, several genes have been identified as causative of AD, including PSEN1, PSEN2 and APP. These genes are involved in the amyloid-beta pathway, and are included in the new diagnostic guidelines for AD as probable AD dementia carriers of a causative AD genetic mutation. In addition, APOE also points to an association with late-onset AD, in that APOE4 allele frequency is high in sufferers. Until recently, we screened these genes using Sanger sequencing. However, it was time consuming and costly if used for routine diagnostic purposes. Over the past few years, high-throughput genome technologies have changed the genetic landscape of AD. Here, we used Ion sequencing technology, which uses an integrated semiconductor device, detects the hydrogen ions at non-optical, high speed. To reveal the causative gene of AD, we investigated 45 Japanese AD patients with positive family histories, and 27 sporadic patients with early-onset AD. All patients were 65 years old or older. We found no variations in 4 genes. Two variations were detected in the PSEN1 gene. One is a known heterozygous missense mutation in the PSEN1 gene (c.488A>G, p.H163R, rs63750598). Another novel heterozygous missense variation in the PSEN1 gene (c.1158C>A, p.F386L, rs429358) was detected. The frequency of patients who had the APOE4 allele was 31% in the familial group. None of the patients with PSEN1, PSEN2 and APP mutations carried the APOE4 allele. In the early onset group, only a novel heterozygous missense variation (c.1262C>T, p.1421M) was detected in the PSEN2 gene. We detected mutations of the PSEN1 and PSEN2 that are causes of AD, using Ion sequencing technology. This technology can help to diagnose autosomal dominant AD that has been difficult to investigate with conventional methods, and will help to expand both clinical knowledge of the disease and consultations with patients and their families.

1340W
Exome sequencing of Parkinson’s disease in order to identify genetic variants with high disease-risk. W. Satake1, Y. Suzuki2, Y. Ando1, H. Tomiyama3, M. Yamamoto4, M. Murata5, N. Hattori1, S. Tsubi6, S. Sugano2, T. Toda7. 1) Div. of Neurol/Mol Bra Sci, Kobe Univ Grad Sch Med, Kobe, Japan; 2) Dept of Med Genome Sci, the Univ of Tokyo, Japan; 3) Dept of Neurol, Juntendo Univ Sch of Med, Tokyo, Japan; 4) Dept of Neurology, National Cerebral and Cardiovascular Center, Osaka, Japan; 5) Dept of Neurol, Nat Cen Hosp of Neurology and Psychiatry, Kodaira, Japan; 6) Dept of Neurol, the Univ of Tokyo, Tokyo, Japan.

Parkinson’s disease (PD) is one of the most common neurodegenerative diseases worldwide, mainly manifesting motor impairment due to degeneration of dopaminergic neurons. Common form of PD appears following a multi-factorial inheritance pattern; that is, the majority of PD-patients show sporadic onset, and 5-10% of the patients show familial aggregation and an elevated relative risk ratio typical of disorders with complex inheritance. Relatives of an affected individual are more likely to have disease-predisposing alleles in common with the affected person than are unrelated individuals. It is reasonable that more genetic variants with high disease-risks of poly- genic PD are thought to exist in patients with family history than in sporadic cases, although there are some cases who will be due to single-gene mutation causing parkinsonism of mendelian inheritance that is masked by small family sizes and incomplete penetrance. Therefore, performing exome sequencing of PD-patients with family history will be effective in order to identify genetic variants with high PD-risk, such as rare variants. We performed exome sequencing using 66 patients from 33 families. We obtained genome of 2 affected individuals from each family. We extracted exome using Agilent SureSelect and massively parallel sequencing using HiSeq2000 (average depth x102). After BWA mapping, GATK calling, and dbSNP132 filtering, we detected an average of 565 nonsynonymous SNVs. Because 2 patients from the same family would share alleles with disease risk, we extracted and detected an average of 244 SNVs (per family) which are sheared between 2 patients from the same family. Among these, 793 SNVs were shared between more than 2 families in total, and 35 SNVs showed variant-frequency <0.25% in data of control samples, which may correspond to candidates of rare variants with high PD-risk. We will examine association between these variants and PD by association studies using exome sequencing of sporadic cases.
1341T
1) Clinical Genomics, Ambyr Genetics, Aliso Viejo, CA, USA; 2) Mayo Clinic, Department of Medical Genetics, Rochester, MN, USA; 3) Children's Hospital & Clinics of Minnesota, St. Paul, MN.

Intellectual disability (ID) is a heterogeneous disorder with a wide phenotypic spectrum. Over 1,700 OMIM genes have been associated with this condition, many of which reside on the X-chromosome. The IQSEC2 gene is located on chromosome Xp11.22 and is known to play a significant role in the maintenance and homeostasis within the neural environment of the brain. Mutations in IQSEC2 have been historically reported as causing non-syndromic X-linked intellectual disability (XLID) characterized by early onset limited intellectual functioning and limited adaptive behavior. Case reports of affected probands show phenotypic overlap with conditions associated with pathogenic MECP2, FOXL1, CDKL5, and MEF2C gene mutations. Affected individuals, however, have also been identified as presenting with additional clinical features including seizures, autistic-behavior, psychiatric problems, and delayed language skills. Although once thought to be a rare cause of XLID, IQSEC2 mutations are becoming more frequently detected through Next Generation Sequencing technologies utilized in clinical diagnosis. To date, a total of four unrelated families and 32 male probands are reported to carry mutations in this gene. Here we report two novel IQSEC2 de novo truncating mutations (c.2582G>C; p.S861T affecting splicing and c.2052_2053delCG; p.C684X) identified through diagnostic exome sequencing, we have so far identified a novel frameshift nonsense mutation in the CC2D1A gene in one family with 3 affected brothers with another frameshift nonsense mutation in the VPS13B gene in one family with 3 affected siblings/first-cousins per pedigree with variable intellectual disability (ID) and autism spectrum disorder (ASD), which fits with a recessive inheritance pattern. We have thoroughly investigated all the symptomatic individuals, including standard clinical examinations, comprehensive psychological evaluations, routine laboratory tests, MRI and EEG investigations; and have collected blood samples from all the affected individuals, their unaffected parents and siblings. We have carried out homozygosity mapping in these 6 ASD/ID pedigrees using high density DNA microarrays. Due to the extreme genetic heterogeneity of ASD/ID, as well as the efficiency and reduced cost of whole exome sequencing, we also performed whole exome sequencing in 2-3 selected affected individuals/pedigree. Results: We have identified extensive runs of homozygosity region in each individual genome, as well as shared identical-by-descent regions among the affected individuals in each or branch of these 6 pedigrees independently. Combined with whole exome sequencing, we have so far identified a novel frameshift nonsense mutation in the CC2D1A gene in one family with 5 affected brothers with intellectual disability and autism behaviors. Genetic validation of frameshift causing mutations in other 4 pedigrees is underway. Discussion: Our results strongly indicate extensive genetic heterogeneity in these consanguineous pedigrees with ASD/ID phenotypes. Direct whole exome sequencing could be the most useful and efficient approach to identify causative disease mutations and to facilitate clinical diagnosis, as well as for much needed early intervention and genetic counseling.
Mutations in the mitochondrial chaperone TRAP1 are associated with the triad of chronic fatigue, pain and gut dysmotility: Crazy, criminal, or just caught in the TRAP? R.G. Bories, L.R. Susswein, S.A. Wong, K.J. McKeran, A.S. Zare, T.R. Foss, H.A. Hornung, K. M. Sheldon. 1) Courtagen Life Sciences, Inc., Woburn, MA; 2) Division of Medical Genetics, Children's Hospital Los Angeles, Los Angeles, CA; 3) Medical School, University of Southern California, Los Angeles, CA.

Functional symptoms, such as chronic fatigue, pain, and gastrointestinal dysmotility, are common and can dramatically impair a person's quality of life, including leading to disability. Patients with these subjective symptoms are often discounted by the medical establishment as being complainers, mentally ill, or even causing the condition (Munchausen, including by proxy). While the etiology of functional symptomatology is largely unknown, these conditions are common in patients with mitochondrial disease, suggesting that genes involved in energy metabolism are good candidates in functional disorder pathogenesis. DNA from 2,700 patients referred for clinical testing was sequenced using Courtagen Life Science's nucSEEK™ Next Generation sequencing platform. The test sequences 1,195 nuclear-encoded genes for proteins that localize to the mitochondria or are associated with phenotypes which mimic mitochondrial disease. Ten patients were found to harbor mutations in the ATPase domain (253V x 7 patients, E192K x 2, E216x 1) of TNF receptor-associated protein 1 (TRAP1). Nine of 10 patients have a triad of chronic pain, fatigue, and gastrointestinal (GI) dysmotility (the 10th patient does not report dysmotility), versus only 10 of 95 patients without TRAP1 mutations referred to us for sequencing (P=3x10−10). Only 1 of 17 patients with TRAP1 mutations outside the ATPase domain have this triad (P=0.0001 v. in domain mutations); rather, these patients present with varying symptoms, and the variants are likely unrelated to their disease. TRAP1 is a mitochondrial antioxidant chaperone involved in preventing apoptosis under oxidative stress. Antioxidant supplementation is often used in the treatment of functional disorders. Therefore, we hypothesize that mutations in the ATPase domain of TRAP1 prevent appropriate antioxidant chaperoning, which predisposes towards the development of functional disease symptomatology. Anecdotal clinical reports suggest substantial improvement in fatigue and pain, among the 4 of our 10 patients known to be on antioxidant therapy. This triad of chronic pain, fatigue, and gut dysmotility is underway, with favorable initial responses. We propose the name TRAP1-Related Disease (TiReD) for this likely-treatable novel condition, and suggest that genes involved in antioxidant defense be investigated more thoroughly as candidates in functional disease pathogenesis.
1345F Novel dominant associations with PANS, autism, anxiety, pain, fatigue and GI dysmotility identified by NextGen sequencing the 1,100 MitoCarta genes in 270 probands. L.R. Sussewitz1, K.M. Shelton1, T. Foss1, A. Zare1, K.J. McKean1, S.A. Wong2, R.R. Trifiletti2, R.G. Boles1, 1) Courtagen Life Sciences, Inc., Woburn, MA; 2) Private practice, Ramsey, NJ.

Often referred to as `functional', `psychosomatic' or `non-organic', disease composed of subjective symptomatology is a frequent cause of disability and poses a major challenge for physicians and researchers. Since these conditions are very common among patients with mitochondrial disorders, mitochondrial genes are good candidates in functional disease pathogenesis. Courtagen has NextGen sequenced all ~1100 genes encoding mitochondrial proteins (MitoCarta) in 270 unrelated individuals referred for clinical testing because of a suspicion of possible mitochondrial disease. Herein, we present novel disease associations. To limit type II errors due to multiple comparisons, candidates are identified based on family studies or an increased prevalence of deleterious-predicted variants among our patients in comparison to controls, and candidates are assigned based on the presence of a unique associated phenotype versus in a group representative of all patients referred for our testing. Below are three among many candidates.

FFPS: This gene encodes ffolypolyglutamate synthetase, an enzyme that has a central role in folate homeostasis. R466C is the only suspicious variant in a mother and three affected children with Pediatric Autoimmune Neurological Syndromes (PANS), and is found in 8/270 (3%) probands v. 0.5% in 1000 Genomes (p=0.0018). 7/8 probands share a PANS-like phenotype defined by the presence of autism, OCD or tics, v. 24/99 in our referral group (p=0.02). ACG86: This gene encodes a protein that transports porphyrazines across the inner mitochondrial membrane. A deleterious-predicted variant was identified in 13 probands v. in 0/41 negative controls (p=0.02). Ten of the 13 probands (77%) were reported to have a PANS-like phenotype including autism, tics, OCD and/or loss of milestones v. 35/100 of our referral control group (p=0.006). The 3 remaining probands have development delay, whereas this phenotype may be unrecognized. TRAP1: This gene encodes a mitochondrial chaperone. In 10 probands, three deleterious-predicted variants in the ATPase domain were detected, demonstrating a clinical triad of chronic fatigue, pain and GI dysmotility (p=3x10^-7). Conclusion: MitoCarta gene identification has revealed a number of new potential disease associations.

1346W Major Depression Susceptibility Loci Identified Using Whole Genome Sequencing in Extended Pedigrees. E.E.M. Knowles1, M.A.A. De Almeida1, J.W. Kent2, T.D. Dyer1, J.E. Curran2, H.H. Goring3, M.A. Carlsson1, R.L. Overa1, D.R. McKay1, E. Sprooten1, J. Blangero1, D.C. Glahn1, 1) Department of Psychiatry, Yale University, New Haven, CT; 2) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 3) Department of Psychiatry, University of Texas Health Science Center, San Antonio, TX.

Background MDD is a common and costly disorder. While effective treatments exist many patients fail to receive them. Identifying genetic markers for depression may provide a reliable indicator of depression risk, which would substantially improve detection, and in so doing enable earlier more effective treatment. The aim of this study was to identify genes for depression modelled as a continuous trait, using whole genome sequence data. The sample comprised 530 Mexican-American individuals from extended pedigrees. Whole genome sequence (WGS) data with a 60-fold coverage were available for the entire sample amounting to ~3.4M SNPs per individual. Association testing was performed using a unitary factor score derived by applying CFA to all items from the Mini-International Neuropsychiatric Interview (M-I-NIPI) and the Mini-Mental State Examination (MMSE) of the Mini-International Neuropsychiatric Interview. Genome wide-significant hits were followed up using gene-specific analyses. WGS analysis revealed two variants on chromosome 3 that were significantly associated with depression, plus a number of variants that reached a suggestive level of significance. The first was located at ~188.0 Mb (922 = 40.15, p = 2.35 x 10^-10) and the other at ~67.0 Mb (922 = 35.21, p = 2.96 x 10^-09). Post-hoc analysis revealed a number of interesting candidate genes. Using a continuous measure of depression combined with WGS data we have identified genes for depression on chromosome 3 and overlap, in part, with identifications from linkage studies and also from the GWAS for depression carried out by the PGC.


Genetic studies have shown a strong association between autism spectrum disorders (ASD) and both inherited and de novo copy number variation (CNV). Studies to date focused on CNV that were >30kb. In the current study, we investigated the role of smaller exome-targeted CNV.

Methods: Whole exome sequencing (WES) was performed on 811 subjects (432 ASD cases, 379 controls) using the Agilent Whole Exome 1.1 kit and Illumina sequencing to generate 75 bp paired-end reads (Neale et al. 2012, Nature, 485:242–5). X-HMM (eXome Hidden Markov Model) Fromer et al. 2012, Am J Hum Genet, 91:597–607 was used to call CNV from exome read depths after removing batch effects. X-HMM calls were filtered to retain CNV with X-HMM quality score > 65, number of exons spanned > 3, CNV length > 1 kb, per-sample number of CNV ≤ 55, per-sample total kb CNV ≤ 18 Mb, and MAF ≤ 1%. The filtered set contains 1386 CNV calls in 559 samples. These results were further stratified on type (deletion/duplication) and CNV size (1–10 kb, 10–30 kb, 30+ kb). Burden analyses were performed with PLINK on each subset to assess the CNV called per sample, the proportion of samples with CNV, and the number of genes hit by CNV per sample. After finding increased burden for all three measures in the 1–30 kb deletions, we chose a subset of those to validate using pQPCR and/or Sanger sequencing.

Results: Strong enrichment was found in the 1-30 kb deletion subset, as measured by the number of CNV calls per sample (p=0.0037), proportion of samples with CNV (28% in cases vs. 21% in controls, p=0.017), and number of genes hit by CNV per sample (p=0.041). We used qPCR to attempt to validate 66 of the 219 deletions in the 1-30 kb range. Of these, qPCR confirmed 55 deletions overlapping the deletion predicted by X-HMM. Further validation of 5 deletions via Sanger sequencing validated three deletions, with two technical failures.

Conclusion: We found a significant (p=0.017) burden of small (1–30 kb) deletions in ASD cases that represents a genetic finding that would be made in 7% of individuals with ASD.

1348F Contribution of LIN7A to developmental disorder. A. Matsumoto1, M. Mizuno1, N. Hamada2, E. Jimbo2, K. Kojima2, M.Y. Momoi3, K. Nagata2, T. Yamagata1, 1) Pediatrics, Jichi Medical University, Shimotsuke, Tochigi, Japan; 2) Department of Molecular Neurobiology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan.

Disorder of synaptic proteins was a major pathophysiology of autism spectrum disorder (ASD). Among them, deletion, duplication and base substitution of SHANK3, one of the scaffolding protein working in the synapse, were detected as causes of autism. We focused on LIN7A and LIN7B that were another scaffolding proteins, and analyzed their contribution to intellectual disability (ID) and ASD. (Patients) Patient 1 is one-year-old boy with ID, spastic diplegia, facial anomalies and thin corpus callosum. Patient 2 is a boy with Asperger syndrome. Patient 3 was a girl with ASD. (Materials and Methods) Their DNA was extracted from lymphocytes after obtaining informed consent from their parents. Array CGH analysis was performed using Agilent Human genome CGH 180K (CNV). Studies to date focused on CNV that were >30kb. In the current study, we investigated the role of smaller exome-targeted CNV.

Conclusion: We found a significant (p=0.017) burden of small (1–30 kb) deletions in ASD cases that represents a genetic finding that would be made in 7% of individuals with ASD.
Cross-disorder copy number variation analysis of Tourette syndrome and obsessive-compulsive disorder. J.M. Scharf on behalf of the TSA/ICG and I2OCFCC. Psychiatric Neurodevelopmental Genetics Unit, Massachusetts Gen Hosp, Boston, MA.

Background: Tourette syndrome (TS) and Obsessive-compulsive disorder (OCD) are highly heritable neuropsychiatric disorders with evidence for shared genetic factors from twin and family studies. TS/OCD segregation patterns also suggest the presence of within-family pleiotropy in the expression of the disorders. In general, genetic pleiotropy has been noted in the copy number variation (CNV) literature, where several large genomic regions have been identified repeatedly to be associated with multiple neurodevelopmental disorders, such as autism, epilepsy, schizophrenia and intellectual disability (ID). This study, which represents the first genome-wide analysis of CNVs in OCD and the largest analysis to date in TS, addressed whether large (>500kb), rare (<1%) CNVs contribute to the genetic architecture of TS and OCD. Methods: Given prior evidence favoring a shared genetic relationship between TS and OCD, a cross-disorder design was employed to maximize power to detect rare events. The sample consisted of 2699 cases (1086 TS, 1613 OCD) and 1789 controls, including a subset of 348 OCD cases recruited as parent-proband trios to allow a de novo CNV analysis. CNVs were called with two algorithms (PennCNV and iPatterm) to ensure reliability. Results: There was no increased burden of large, rare CNVs in the cross-disorder analysis or in secondary, disorder-specific analyses. However, a trend (p=0.06) was noted for a 3.5-fold increase in deletion burden in regions previously associated with other neurodevelopmental disorders. Further examination revealed that a single locus, 16p13.11, harbored most of the neurodevelopmental burden (5 case deletions: 0 control deletions, p=0.09 in current study, p=0.025 compared to published control rates). Furthermore, three of these 16p13.11 deletions were confirmed as de novo, supporting the etiological significance of this region. Discussion: These results demonstrate that TS and OCD are associated with known neurodevelopmental deletions, most notably at 16p13.11 which previously was associated with ID and autism. 4/5 cases with a 16p13.11 deletion had OCD (3 OCD only, 1 OCD + tics) whereas one case had TS only; none had ID or ASD. These data are consistent with pleiotropy at the 16p13.11 locus with TS and OCD probably sharing the same biological networks involved in other neurodevelopmental disorders. CNVs in OCD and TS will be important to expand the range of phenotypes associated with CNVs implicated across neurodevelopmental disorders.

Large deletion of the PRRT2 gene frequently involves paroxysmal dyskinesia in Korea. Sj. Lee1, MW. Seong2, HJ. Kim3, S. Kim4, SS. Park5, BS. Jeon2. 1) Department of Laboratory Medicine, College of Medicine, Seoul National University Hospital, Seoul, Korea; 2) Department of Neurology and Movement Disorder Center, Parkinson's disease study group, and Neurosciences Research Institute, College of Medicine, Seoul National University, Seoul, Korea.

Introduction: Recently, it has been found that mutations of PRRT2 cause paroxysmal kinesigenic dyskinesia (PKD) and other paroxysmal dyskinesias (PxDs). However, detailed clinical features of the patients with PRRT2 mutation in comparison with those without mutation are not well described. Furthermore, 16p11.2 microdeletions including PRRT2 also have been reported in patients with PKD but it is unknown to what extent PRRT2 deletion contributes to development of PKD and other PxDs. Methods: To address these issues, we performed mutation screening in 30 Korean patients with PxDs by sequence analysis and gene dosage analysis of PRRT2, then analyzed their clinical features. Results: Overall, genetic abnormality in PRRT2 was identified in 9 patient (30%), 3 of the 7 familial cases (43%) and 6 of the 23 sporadic cases (26%). The previously reported c.649dupC and c.649delC were found in 5 and 1 patient, respectively, and a novel mutation c.323_324delCA was found in 1 patient. The other two patients were found to have PRRT2 deletion involving at least exon 2 and 3. Compared with mutation-negative cases, age at PxDs onset was earlier in mutation-positive cases. However, there was no difference in other clinical features. Contrary to common belief that patients with PKD show excellent response to carbamazepine, 4 mutation-positive patients taking carbamazepine reported only partial responses. Conclusion: Our result shows that PRRT2 is the common causative gene for PxDs in Korea. The PRRT2 deletion might frequently involve development of PxDs and gene dosage analysis should be included in the genetic test for PxDs. The mutation detection rate in familial cases is lower than expected and compared with those in previous studies. This result suggests that other genetic, epigenetic or environmental factors are related to paroxysmal dyskinesia. Further studies are needed to identify the possible effect of these factors.

1) Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas, Campinas, Brazil; 2) Porto Alegre Epilepsy Surgery Program, Hospital São Lucas, PUCRS, Porto Alegre, Brazil; 3) Clinical Hospital of Faculty of Medicine, USP, Sao Paulo, Brazil; 4) Department of Neurology, Faculty of Medical Sciences, University of Campinas, Campinas, Brazil.

Purpose

Although several genes have been implicated with increased susceptibility to epilepsy, only a few are considered potentially useful for clinical testing. SCN1A mutations are associated with phenotypes within the spectrum of generalized epilepsy with febrile seizures plus (GEFS+), but the prognostic value of these mutations and a possible correlation with the different clinical subtypes remain unclear. Therefore, the aim of this study was to expand the knowledge on the clinical use of genetic testing for the most severe phenotypes within the GEFS+ spectrum, namely Dravet and Doose syndromes.

Methods

SCN1A mutation screening was performed in 21 patients with Dravet and 15 with Doose syndrome. Potentially deleterious mutations found were investigated in 100 individuals without epilepsy. In addition, eight algorithms were used to analyze the possible impact of missense mutations in protein function. Furthermore, MLPA analyzes were performed to detect copy number variations within SCN1A. Results

We identified 14 potentially deleterious mutations in 15 patients with Dravet: six missense, four splice-site, two frameshift and two in-frame deletions. None of these were found in controls. Furthermore, all missense mutations are predicted to affect protein function. Two patients are monozygotic twins and share an identical mutation, an 18-basepair deletion, which was detected by both sequencing and MLPA. Interestingly, they also presented a missense mutation, which differently from the six mentioned above, was predicted as benign by five algorithms. In addition, this missense change was also present in their unaffected mother and in a control subject, strongly suggesting that the main cause of epilepsy in these patients is the deletion rather than the missense mutation. In contrast, only one patient referred to us as presenting Doose showed a potentially deleterious mutation, although this patient may have indeed a borderline/atypical phenotype. Patients with SCN1A Dravet syndrome showed a high frequency of SCN1A mutations (71%). Most mutations are missense (43%) and located in functionally important regions. However, we found that SCN1A testing in patients with classic Dravet syndrome does not seem to be clinically relevant. In addition, we achieved a good discrimination power for identifying potential deleterious sequence variations in SCN1A, including missense changes by incorporating prediction algorithms and sequencing of control individuals.

Linkage analysis and exome sequencing in a large highly inbred consanguineous kindred to identify idioopathic generalized epilepsy genes. F. Tuncer1, S.A. Uqur Iseri2, M. Caik3, A.O. Caglayan2, A. Iscan4, M. Gunel2, O. Ozbek5.

1) Genetics, Istanbul University Institute for Experimental Medicine, Istanbul, Turkey; 2) Department of Pediatric Neurology, Harran University, Faculty of Medicine, Sanliurfa, Turkey; 3) Departments of Neurosurgery and Genetics, Yale School of Medicine, New Haven, CT, USA; 4) Department of Pediatric Neurology, Bezmialem Vakif University, Faculty of Medicine, Istanbul, Turkey.

Epilepsy is a complex neurological disorder affecting 1% of the world’s population. Among different types of epilepsies, idiopathic generalized epilepsies (IGEs) are characterized by bilateral and synchronous generalized seizures in the absence of detectable brain lesions or metabolic abnormalities. Thus, the primary etiology for this disorder is believed to be genetic. The present study includes a large highly inbred consanguineous kindred with multiple IGE affected individuals and an ultimate aim in identifying novel epilepsy gene(s) to delineate the molecular basis of this disorder. Physical, neurological and electroencephalography (EEG) examinations were performed on the subjects recruited with information on family history, revealing 6 affected family members composed of 4 siblings and 2 cousins presenting varying seizure types. Experimental approaches included SNP genotyping of affected and 4 unaffected family members using Illumina Human HumanCytoSNP-12 BeadChip (300K). Genotype data was utilized using easyLinkage Plus software platform, where Mendelian genotyping errors were determined using PedCheck, multipoint lod scores were calculated under the assumption of autosomal recessive inheritance and haplotypes were constructed through GeneHunter. Linkage analyses were performed utilizing 4 affected sibs with their healthy brother and parents. Linkage peaks obtained were investigated through homozygosity mapping, which revealed partitioned homozygous regions in the affected cousins that were homoygous in the affected sibs. To facilitate in finding epilepsy genes, Illumina HiSeq2000 was used for exome sequencing of the affected child. The candidate gene described to have drug resistance seizures and mental retardation. This child was chosen as the representative phenotype among the affected family members, due to the severest clinical presentation. Exome data obtained from this child was analyzed utilizing the pipeline developed by Yale University’s bioinformatics team. Data was filtered for homoygous variants that either reside under the linkage peaks, within the homoygous regions shared by affected members and/or are novel and not caught by previous analyses. The candidate gene was also investigated via Sanger sequencing, followed by familial segregation underline the current experiments. Prospective work includes healthy population screening for the candidate variant that shows familial segregation in line with affection status.
"Game of Exomes" and Autism Spectrum Disorder. M. Cuccaro, N. Dueker1, E.R. Martin1, A.J. Griswold1, H.N. Cukier1, S. Sillier1, J. Jaworski1, I. Konidari1, P.L. Whitehead1, M. Schmidt1, J.R. Gilbert2, J.L. Haines1, M.A. Pericak-Vance1, 2. 1) Hussman Institute for Human Genomics, Univ Miami Sch Med, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Vanderbilt University School of Medicine, Vanderbilt University, Nashville, TN.

Autism spectrum disorder (ASD) is a highly prevalent developmental disorder, affecting an estimated 1 in 88 individuals, and is associated with significant morbidity. Despite demonstrating high heritability estimates, only a small proportion of risk is explained by single nucleotide polymorphisms (SNPs). Recent research, including genome wide association studies, has focused on the effects of common variants, rare variants and copy number variants on the risk of ASD. Previous research has identified suggestive evidence for a role of RVs in ASD risk. Therefore, we analyzed RV data to identify loci associated with ASD. Participants were drawn from a large, family-based study of ASD and included 995 unrelated cases and 650 controls that were genotyped using the Illumina HumanExome-12v1 Array (HIHG sample). In addition, whole exome sequencing data on 1,514 cases and 866 controls from the ARRA autism sequencing consortium was utilized for replication and joint analyses (ARRA sample). To identify individual SNPs associated with ASD, we performed Fisher's Exact Test on all SNPs with MAF<5% and Fisher's method for combining p-values to perform meta-analyses. To identify genes associated with ASD we performed gene-based analyses testing autosomal genes genotyped on the Array for association with ASD. The sequence kernel association optimal test (SKAT-O), Cochran-Armitage (CA) Sum and CA Max tests were used for these analyses. A total of 108,045 SNPs were included in single-SNP analyses and 462 sub-haplotypes, as well as wide-coverage exome analysis, and 17 were associated with p<0.01 in at least one test. We also identified 19 ASD candidate genes with p<0.001. The most significantly associated being adenylate cyclase 5 (ADCY5) (SKAT-O p=0.002, CA Sum test p=1.8x10⁻⁴, CA Max test p=3.9x10⁻⁴). ADCY5 is an excellent ASD candidate as it plays an important role in neurotransmitter signaling and has been previously associated with de novo mutations in this gene. While our study failed to identify RVs and symptomatic therapies.

The latest sequencing and genotyping technologies to comprehensively assess joint effects of variants with MAF<1%. We will present results using our independent case-control series (747 progressive supranuclear palsy and thus identify novel targets for both neuroprotective and symptomatic therapies.

Role of MAPT variation in neurodegenerative disorders. C. Labbé1, A. Ortola1, S. Rayaprolu1, R. Uitti2, D. Dickson2, Z. Viszoke2, O. Ross1, 1) Department of Neurosurgery, Mayo Clinic, Jacksonville, FL; 2) Department of Neurology, Mayo Clinic, Jacksonville, Florida, USA.

Tou inclusion defines the neurodegenerative diseases called tauopathies. Rare variants in MAPT, the gene encoding protein tau, cause tau dysfunction leading to neurodegeneration. A common non-combining MAPT haplotype (MAPT H1) has been associated to several tauopathies including progressive supranuclear palsy, corticobasal degeneration, and Pick’s disease; yet MAPT H1 has also been implicated in the risk to Parkinson’s disease (which is not a traditional tauopathy). Preliminary sub-haplotype analyses suggest that different genetic variants on the MAPT H1 haplotype associate with each of these disorders. To date it remains unclear which variant(s) at the MAPT locus is(are) responsible for the risk and what is the underlying pathomechanism of disease. The objective of this study is to identify causal variations for Parkinson’s disease and progressive supranuclear palsy (and other tauopathies) within the MAPT region using next generation sequencing technologies. We captured the entire MAPT gene and 10kb on each side, a 154kb genomic region that was sufficiently unbiased in our HiHG sample. 653 patients with progressive supranuclear palsy, 20 patients with Pick’s disease, 30 patients with corticobasal degeneration and 300 controls using a pooling strategy (10 DNA samples/pool). We used the Haloplex system and designed 4249 amplicons for a total coverage of 96.3%. We selected: (a) common variants tagging each of the MAPT sub-haplotypes (frequencies >5%); and (b) rare variants most likely to be functionally relevant. Common and rare genetic variants identified are being genotyped in our independent case-control series (747 progressive supranuclear palsy patients, 1356 with Parkinson’s disease, 235 clinically diagnosed patients vs 689 controls), with linear regression analyses to study all variants with a minor allele frequency (MAF) >1%. A collapsed marker approach will assess joint effects of variants with MAF<1%. We will present results using the latest sequencing technologies. The study should be interpreted as a pilot study but the significant differences will be the result of sequencing errors and confirmation of variants of interest will serve as the next step of this analysis. The results support our strategy and identify patient specific genetic changes that may lead to schizophrenia. The latest sequencing technologies re-enforce the complex role of genetic variability, some inherited and others acquired during parental meiosis and/or mitosis during ontogeny. Even monozygotic twins are not identical and each individual may be a mosaic. This is supported by a high mutation rate and the persistence of genetic diseases with a severely reduced fecundity in human populations.
Whole genome DNA sequencing identifies variants showing allelic association with bipolar affective disorder. A. Fiorentino, N. O’Brien, A. McQuillin, A. Narula, A. Angrist, R. Kasowski, R. B this study, we have sought to detect these variants using whole genome sequencing data from 99 BD subjects. Two of the best implicated BD susceptibility genes, ANK3 and CACNA1C, were included in the analysis. We also sequenced the two new possible allelic base pair changes that we have found.


Whole genome sequencing data from 99 BD subjects. Two of the best implicated BD susceptibility genes, ANK3 and CACNA1C, were included in the analysis. We also sequenced the two new possible allelic base pair changes that we have found.


The majority of human T-cell leukaemia virus type-1 (HTLV-1)-infected individuals remain lifelong asymptomatic carriers, but some infected individuals develop HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP). HAM/TSP is a chronic inflammatory disease of the central nervous system that presents as slowly progressive spastic paraparesis with neurogenic bladder. Previous studies revealed that several host genetic factors were associated with the development of HAM/TSP. To identify the genes that contribute to HAM/TSP, we analyzed familial HAM/TSP. We performed whole-exome sequencing in 32 patients with familial HAM/TSP, 20 patients with sporadic HAM/TSP, and 20 HAM/TSP carriers. Rare variants with minor allele frequencies of less than 5% in 1000 genomes were considered for analysis. To predict loss-of-function genes, we determined whether missense mutations predicted by PolyPhen2 were damaged or frameshift indels. A total of 28,052 SNPs and 2097 indels were identified among patients with familial HAM/TSP. Rare variants identified in patients with familial HAM/TSP but not in HAM/TSP carriers were analyzed. The maximum number of patients harboring each variant was 11. The affected rare variants shared by a minimum of 5 patients were investigated, and refining the variants to those associated with a predicted loss of function resulted in the selection of 23 genes. These genes were related to cell communications, and 1 gene was related to the nervous system. Some variants were detected in patients with sporadic HAM/TSP. These results demonstrate that rare variants are related to the development of HAM/TSP. Whole-genome sequencing is a useful method to identify rare variants in common diseases.
1364W
Discovery and Analysis of Rare Variants for Bipolar Disorder by Exome Sequencing in Multiplex Families. S. Ramdas1, A.B. Ozel2, J. Li1. 1) Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Bipolar disorder (BPD) is a severe psychiatric disease marked by alternating manic and depressive episodes. Twin and family studies have shown that BPD is highly heritable; but exome and genome-wide association studies (GWAS) remains a major barrier to identifying its genetic basis. In this study we focus on multi-generation multiplex BPD families in the NIH repository. We hypothesize that these families manifest a nearly Mendelian or at least highly penetrant subset of bipolar cases and may transmit one or more high-impact coding variants that alter the function of key neurodevelopmental or neural signaling genes. We genotyped and performed exome sequencing for 82 affected individuals in 29 families representing first cousin pairs or more distant relatives, aiming to detect functionally damaging rare variants in regions shared by affected relatives. Comparison of exome sequencing and HumanExome Beadchip genotype data revealed high levels of concordance: on variant sites that were both genotyped and had a called genotype in the exome data, overall concordance is 0.98 and heterozygote concordance is 0.91. Variant filtering were applied to identify those that are bi-allelic, have a minor allele frequency of <5% in European samples from the Exome Sequencing Project, and are damaging missense, nonsense, or splicing site variants, resulting in a median of 336 variants per family. As Exome chip genotype data allow inference of genomic segments shared among family members identically by descent (IBD), we used Beagle to identify IBD regions shared by BPD cases within each family. This led to a further reduction in the number of candidate variants. In one family, only one gene, FOXO4 (forkhead box protein D4), was identified after filtering. This gene has previously been implicated in suicidality and Obsessive Compulsive Disorder (OCD). Members in another family carry damaging variants in DRD5 (Dopamine receptor D5) and GRIK3 (Glutamate receptor, ionotropic kainate 3), which were reported in association studies of bipolar and schizophrenic phenotypes respectively. Additional filtering and validation are underway for the candidate variants across all families. (This study is supported by the IMHRO - Johnson & Johnson Rising Star Translational Research Award.)

1365T
Gene based analysis on major depressive disorder using 12,000 case control samples sequenced at one fold coverage. Y. Li1, J. Wang1, S. Shi2, Y. Chen3, J. Marchini4, K. Kendler1, J. Flint1 on behalf of the CONVERGE Consortium. 1) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, Oxfordshire, United Kingdom; 2) Beijing Genomics Institute, Floor 9 Complex Building, Beishan Industrial Zone, Yantian District, Shenzhen 518083, P.R.China; 3) Huashan Hospital of Fudan University, No.12 Middle Wulumuqi Road, Shanghai, P.R.China; 4) CTSU, Richard Doll Building, Old Road Campus, University of Oxford, Headington, Oxford OX3 7LF, United Kington; 5) Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, PO Box 980128 Richmond VA 23298, USA.

We present results for low pass sequencing from the CONVERGE project, a case-control study of major depressive disorder (MDD) using 12,000 Han Chinese women. To date, no genetic loci have been robustly associated with MDD using genome-wide association based on genotypes called from arrays. Low pass sequencing allows novel approaches to genotyping, particularly in accessing low frequency variants. We applied gene enrichment tests on functional SNPs with different levels of filtering and allele frequency cut-offs to test whether cases were enriched with disease causal SNPs than controls in any genes, and if so, where they are on the allele frequency spectrum. The filtering includes base quality, mapping quality, exclusion of non-pair end reads and alternative allele supported by at least two reads. The gene enrichment analysis has given us candidate genes, which can now be verified by targeted re-sequencing. Furthermore we report higher copy number of the mitochondrial genome in cases with MDD. This is the largest GWAS using whole genome low coverage sequencing data. Although we face the challenge of calling and imputing low frequency variants accurately, our pipeline allows different levels of filtering at an individual level, which enables us to examine rare variants aggregate down to a singleton level. The results help to answer an important question about whether rare functional variants are responsible for the genetic susceptibility to MDD.

1366F
Exome sequencing identifies de novo missense mutation in KCND2 in identical twins with autism and seizures that results in slow potassium channel inactivation. S. Nelson1,2, H. Lee2, M. Lin2, H. Kornblum4, D. Papazian3, J. Mott1,1, J. Flint1,1, K. Kendler5,1, R. Flint1,1, J. Marchini4, K. Kendler1, J. Flint1, on behalf of the CONVERGE project. 1) Dept Human Genetics, UCLA Medical Ctr, Los Angeles, CA; 2) Dept Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 3) Dept Physiology, UCLA, Los Angeles, CA; 4) Dept Pediatrics, Los Angeles, CA.

There have been numerous studies and case reports showing co-existence of autism and epilepsy suggesting significant overlap between the two phenotypes. However, the relationship between the two on the molecular level remains unclear. Here, we performed whole exome sequencing on a family with identical twins affected with autism and severe, intractable seizures. A de novo variant p.Val404Met was identified in KCND2 gene that encodes A-type potassium channel Kv4.2 which forms the pore of the somatodendritic subthreshold A-type potassium current (I\textsubscript{SA}) channels. The p.Val404Met variant is a novel variant and occurs at a highly conserved residue within the C-terminal end of the transmembrane helix S6 region that makes up the ion permeation pathway. In available exome data, the transmembrane domains in Kv4.2 protein have a low frequency of predicted-to-be damaging variants relative to the cytoplasmic domains indicating strong selection against variation in the pore domains. To investigate the functional effect of the heterozygous p.Val404Met mutation on the properties of Kv4.2 channels, reference and mutant proteins were expressed together and individually in Xenopus oocytes and analyzed using a two electrode voltage clamp. The mutant channel demonstrated significantly reduced rate of both the opening and inactivation of the channel. Co-expression of the reference with the mutant channels was mostly indistinguishable from the mutant channel alone demonstrating gain of function mutation. The effect of the mutation on closed-state inactivation was evident in the presence of the auxiliary subunits that associate with Kv4 subunits to form I\textsubscript{SA} channels in vivo. Our results taken together with previous reports strongly suggest that the size of the residue at position 404 is a major factor in determining the likelihood of inactivation and given the key role of closed-state inactivation in determining the steady state availability of I\textsubscript{SA} channels in neurons, the dominant phenotype of p.Val404Met is likely to disrupt the normal physiological function of I\textsubscript{SA} channels. In the family, the novel de novo variant in KCND2 at a conserved residue, coupled with physiological evidence of the mutant protein disrupting the potassium current inactivation impose a strong indication that KCND2 is the causal gene for the epilepsy in this family, and provides suggestive evidence of a role in the autistic features that they exhibit.
1367W

Amish revisited: A next-generation sequencing study of bipolar disorder among the Plain people. L. Hou¹, N. Akula¹, L. Kassam¹, D. Chen¹, Y. Yao², J.L. Haines³, M. Pericak-Vance⁴, T.G. Schulze¹, F.J. McMahon¹. ¹Human Genetics Branch, National Institute of Mental Health Intramural Research Program, Bethesda, MD USA; ²Unit of Statistical Genomics, National Institute of Mental Health Intramural Research Program, Bethesda, MD, USA; ³Center for Human Genetics Research, Vanderbilt University, Nashville, TN USA; ⁴Hussman Institute for Human Genetics, University of Miami, Miami, FL USA; ⁵University of Goettingen, Goettingen, Germany.

Bipolar disorder (BP) is a common, highly heritable mental illness. A number of risk loci have been identified by GWAS, but the large proportion of unexplained genetic variance suggests that rare variants might contribute to risk of the disorder. Isolated populations with large families, like the Amish, have many advantages for discovering rare genetic susceptibility loci, such as lower allelic and locus heterogeneity, higher environmental and phenotypic homogeneity, and enrichment of rare disease alleles. We are ascertaining distantly related cases of BD and their relatives from Plain (Amish and Mennonite) communities in the US. So far we have 170 study participants, 94 of whom have been assigned a diagnosis of BD with high confidence. Identity-by-descent (IBD) analysis of SNP array (Illumina OmniExpress) genotypes (by use of Beagle v 3.3.2) demonstrates sharing of many long (>5 cm) chromosomal segments, consistent with known pedigree relationships, state of residence, and Amish vs. Mennonite affiliation. To identify IBD segments that were shared more often by case-case pairs than others, we combined the OmniExpress data with SNP array (Affymetrix 6.0) data from 893 Midwestern Amish individuals (collected by JLH & MPV), extracted 154,271 overlapping markers, excluded closely related pairs (π-hat >0.1), and carried out a shared segment analysis with PLINK (v 1.07). Several highly shared IBD segments were identified on chromosomes 1p34.3, 2q31.1, 7p22, 7q36.3, 3p24.2-p24.1, and 10q22.1 at permutation p-values ranging from 0.01 to 2E-05. Interestingly, SNPs in several of these segments have been reported to be associated with bipolar disorder by previous GWAS. Exome sequencing has so far been carried out on 21 distantly related cases using the Illumina HiSeq system. The BWA/GATK package was applied for single nucleotide and insertion-deletion variant calling, with filtering of common SNVs (MAF > 5% in the Exome Sequencing or 1000 Genomes Project). Five novel, nonfunctional SNVs and 9 potentially functional SNVs, including 8 non-synonymous and 1 frame-shift insertion, were found within the highly shared regions and were each detected in at least 4 cases. While we have shared segment analysis and sequencing in larger samples is needed, these preliminary results suggest that rare variants may contribute to BD in Plain populations.

1368T


Speech sound disorders (SSDs) are disorders of the speech output mechanism, often diagnosed before the age of 4. Children with SSD fail to develop speech that is easily understood by others, due to omitted, distorted, or substituted speech sounds. SSD runs in families and there is evidence of a genetic etiology but the genetic pathways are poorly understood. One proposed SSD subtype is childhood apraxia of speech (CAS), a severe SSD subtype that is thought to result from deficits in motor programming. In some forms of syndromic CAS, the genetic etiology is known, e.g., GALT mutations in galactosemia and FOXP2 mutations in severe speech and language disorder. The genetic etiology of idiopathic CAS has not been delineated. We study multigenerational families with familial idiopathic CAS from phenotypic and genetic perspectives. Our results indicate that motor programming deficits are not limited to the speech production system but can also be observed in hand motor tasks, primarily when the task involved alternating-sequential movements, not repetitive ones. An even more comprehensive deficit in sequential processing emerged from group differences in adults with, and without, a childhood history of CAS in tasks involving sequential processing on linguistic and cognitive levels, for instance nonword reading and spelling. To investigate the genetic etiologies of CAS, we use approaches made possible by our multigenerational family design. A combination of linkage analysis, exome sequencing, and identity-by-descent studies are leading to the identification of new candidate regions and genes for CAS. In one family consisting of 23 individuals in three generations, two new candidate genes, expressed in relevant brain regions, were identified on chromosomes 3 and 5. Similar genetic analyses in other multigenerational families are underway. The long-term goal of this study is to create a catalogue of genetic etiologies in SSD to aid in early identification of infants at genetic risk for SSD and development of effective intervention and prevention approaches to avoid or ameliorate the devastating social, psychological, and educational trauma of unintelligible speech in the preschool and early school years. This project is supported by the University of Washington Center for Mendelian Genomics funded by NIH grant 1U54HG006493 to Drs. D. Nickerson, J. Shendure and M. Bamshad.
Comprehensive annotation and analysis of noncoding autism spectrum disorder risk loci by targeted massively parallel sequencing. A. J. Griswold, D. Van Bronswijk, J. M. Jaworski, S. Stift, M. A. Schmidt, W. Hulme, I. Konidari, P. L. Whitehead, J. R. Rantsu, S. M. Williams, R. Menon, M. L. Cuccaro, E. R. Martin, J. L. Haines, J. R. Gilbert, J. P. Hussman, M. A. Pertschuk-Vance, 1, 2 J. P. Hussman Institute for Human Genetics, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL, USA; 3) Vanderbilt University School of Medicine, Vanderbilt University, Nashville, TN, USA; 4) Rollins School of Public Health, Emory University, Atlanta, GA, USA; 5) Human Genetics for Autism Network, Ellicott City, MD, USA; 6) Hussman Foundation, Ellicott City, MD, USA.

Genome-wide association studies (GWAS) of autism spectrum disorder (ASD) show that common variants with small effects do not greatly impact ASD risk. Exome sequencing studies, where implicating rare, de novo protein coding variations in a few genes, suggest that many rare variants in hundreds of genes influence ASD. A largely unstudied hypothesis is that rare, functional variants in non-protein coding regions of the genome contribute to ASD. To identify such variants, we sequenced 17 kb of ASD associated genes and genomic loci (Hussman et al. 2011) in 919 cases and 854 controls. We targeted exons, including UTRs, of 681 genes, and conserved regions in their introns, 5 kb from the 5’ and 3’ ends, and ASD associated intergenic regions. This is the first large scale study of noncoding variants in ASD. We identified 427,676 single nucleotide variants (SNVs), of which only 20,293 were in coding exons. We annotated each noncoding SNV with databases including ENCODE (DNaseI sensitivity and transcription factor binding sites), VISTA (noncoding segments with demonstrated enhancer activity), and GENG (noncoding transcript evidence based noncoding transcript). SNVs without predicted functional effects are assessed using predictions for splice enhancing ability, RNA secondary structure effects, and transcriptional regulatory potential. Restricting our analysis to the 358,432 rare, noncoding SNVs (MAF <0.01 in our dataset), 113,896 are predicted to have a functional role in at least one database. Of these, 11 occur at least once and uniquely in cases and had a score of 2a in the RegulomDB, suggesting they affect a known transcription factor binding site motif. Two of these are in the potassium channel gene KCNJ2, in AR and AUT2 families. SNVs are predicted to have a functional role using the human genome reference sequence (SeqTrimMap and Bowtie), and sequence kernel association test (SKAT) to detect association between ASD and sets of SNVs in noncoding elements while adjusting for population stratification. SKAT did not identify multiple testing corrected association of ASD with any noncoding feature. However, a promising nominal association was detected in enhancer hs1316 (p=0.03) that enhances expression in the midbrain of mouse embryos.

To increase statistical power and identify new variants, sequencing of an additional 1525 cases and 425 controls is underway. This work adds noncoding SNVs to the growing list of ASD risk loci and may implicate biological mechanisms beyond protein coding changes contributing to ASD etiology.

Investigation of the role of microRNAs in autism by deep RNA-sequencing, J. A. Lamb, F. Marriage, E. Tsitsiou, P. Wang, L. Zeef. 1) Centre for Integrated Genomic Medical Research, Faculty of Medical and Human Sciences, University of Manchester, Manchester, United Kingdom; 2) Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom.

Background. Numerous heterogeneous rare intragenic variations and copy number variants have been recently identified in autism, with predicted effects on protein function. Non-coding RNA has an important function in genome regulation, and there is evidence for a role of microRNAs in development and in several neuropsychiatric and neurological disorders. MicroRNAs act as pleiotropic regulators of gene expression, targeting a large number of neuronal genes, suggesting that genetic variation in microRNAs may underlie the clinical heterogeneity of autism. Objectives. To identify novel genetic variation in microRNAs and differentially expressed microRNAs in individuals with autism compared to healthy controls. Methods. Genomewide microRNA expression profiling was carried out on lymphoblastoid cell line RNA from 60 individuals with autism and 10 healthy controls on the SOLID 4.0 analyzer. Sequence data was processed and mapped to the human genome reference sequence (SeqTrimMap and Bowtie), and against all known human mature microRNAs (miRBase v19). Differentially expressed microRNAs were identified using DESeq and false discovery rate correction. Dysregulated microRNAs were validated by qRT-PCR. Variant calling was carried out using Samtools and filters including minimal base and map quality. Results. Seven differentially expressed microRNAs were identified in samples from individuals with autism compared to healthy controls (6 up-regulated, 1 down-regulated; FDR P<0.05). Altered expression of microRNAs was associated with the position of microRNA upstream sequence at 120th and 121st nucleotide, indicative of possible arm switching in individuals with autism. Seven novel sequence variants were identified in the seed region of the mature microRNA in individuals with autism that were not present in control samples or in dbSNP (137). These may alter target mRNA binding and are therefore predicted to be functionally deleterious. Validation of these novel variants and segregation analysis within families of affected individuals is ongoing by Sanger sequencing. The putative mRNA targets of differentially expressed microRNAs and microRNA2 targets are being investigated using in silico and in vitro target prediction programmes (Ingenuity Pathway Analysis). Conclusion. A number of differentially expressed microRNAs and novel variants within microRNA genes have been identified in individuals with autism. These may affect target mRNA binding, suggesting a possible pathogenic role in individuals with autism.
Somatic Instability in Sporadic Amyotrophic Lateral Sclerosis? C.S. Leblond1,2, J.B. Rivière1,2, M.J. Strong2,5, K. Voikening3,6, P. Hince2, D. Spiegelman2, A. Dionne-Laporte3, J. Robertson4, L. Zmiana1, P.A. Dion3, G.A. Rouleau2, 1) McGill university, Montreal, Canada; 2) Montreal Neurological Institute and Hospital, Neurology and Neurosurgery department, McGill University, Montreal (QC) Canada; 3) Equipe Génétique des Anomalies du Développement (EA 4271 GAD), Université de Bourgogne, France; 4) Laboratoire de Génétique Moléculaire, Centre Hospitalier Universitaire Dijon, France; 5) Molecular Brain Research Group, Robarts Research Institute, University of Western Ontario, London (ON) Canada; 6) Department of Clinical Neurological Sciences, Schulich School of Medicine and Dentistry, University of Western Ontario, London (ON) Canada; 7) Tanz Centre for Research in Neurodegenerative Diseases, and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto (ON) Canada; 8) Sunnybrook Health Sciences Centre, Toronto (ON) Canada; 9) Pathology and Cellular Biology department of Montreal University, Montreal (QC) Canada.

Amyotrophic lateral sclerosis (ALS) is a devastating neurological disease characterized by the degeneration of upper and lower motor neurons of the motor cortex, brainstem and spinal cord. The majority of ALS cases are sporadic (sALS) and only 5-10% have a family history (fALS). sALS cases are clinically and pathologically very similar to FALS cases. Whereas 60-80% of fALS cases are explained by germlinal mutations, the genetic etiology of sALS remains unclear. The lack of genetic evidence in sALS suggests that inherited or germlinal de novo mutations are very rare events in this disease. Nonetheless it is important to keep in mind that current genetic evidence for sALS comes from genetic examinations made using blood DNA for the most part and to a lesser extent from available post-mortem neuronal tissues. Here, we hypothesize that somatic mutations appear early during the embryogenesis events that underlie the development of the spinal cord. These somatic events could later trigger the development of ALS in ALS cases, as observed in patients with germine ALS-predisposing gene mutations. Given ALS is a motor neuron disease that essentially affects the corticospinal tract, we are in the process of preparing a total of one thousand sections isolated from flash frozen SALS patient spinal cords. These sections will be individually screened for the presence of mutations in C9ORF72 by repeat-primed PCR (RP-PCR) since repeat mutations have been shown to be unstable during replication and DNA repair. Following this, we will look for point mutations in SOD1, TARDBP and FUS using a targeted deep sequencing method ‘single molecule Molecular Inversion Probe (smMIP)’ specifically designed to identify rare somatic events. The smMIP method has been shown to be highly accurate and sensitive for the detection of sub-clonal variations based on the tagging of amplified DNA molecules that originated from one of the various inputted DNA molecules from the different cells found in a tissue or sample. Preliminary results showed that our RP-PCR conditions are sensitive enough to detect expansions at a low level of mosaicism (20%). To date, we have examined 499 spinal cord sections that were derived from 13 sALS patients for the presence of expanded C9ORF72 alleles but have thus far detected no expansion.

Towards the identification of new genes implicated in recessive early-onset forms of Parkinson’s disease, S. Lesage1, A.S. Cocque1, A.L. Leutenegger1, A. Honoré2, C. Condroyer3, A. Düür1,3, A. Brice1,3, 1) INSERM UMR_S975, ICM, Pitié-Salpêtrière hospital, Paris, France; France; 2) INSERM UMR_S946, Fondation Jean Dausset-CEPH, Paris, France; 3) Département de Génétique et Cytogénétique, Hôpital de la Pitié-Salpêtrière, Paris, France.

BACKGROUND: Parkinson’s disease (PD) is a progressive neurodegenerative disorder, probably resulting from the interplay between genetic and environmental factors. To date, more than 15 loci have been identified and 6 genes are now confirmed to be causative for monogenic forms of PD, including 4 genes (Parkin/PARK2, PINK1/PARK6, DJ-1/PARK7 and ATP13A2/PARK9) that account for only ~50% of PD cases with early-onset (EO) autosomal recessive (AR) inheritance. Hence, a significant proportion of inherited PD cases still remain unexplained genetically. AIM: We propose to use an integrative approach combining homozygosity mapping/genomic rearrangement detection and targeted exome sequencing in consanguineous families with an early age at disease onset (<55 years) to identify new causative genes for recessive PD. PATIENTS AND METHODS: Homozygosity mapping was performed in a series of 160 PD consanguineous families or isolated patients originating from Europe (n=38), North Africa (n=57), Turkey (n=61) and other countries (n=4) excluded for PARK2, PINK1, DJ-1, and the common LRRK2 G2019S mutation, using SNP genotyping microarrays (Illumina Infinium HD HumanCytoSNP-12 BeadChip) and an original linkage statistics program that allows the inclusion of individuals for whom genealogical information is lacking. Using the same SNP microarrays, rare large deletions and duplications were searched for. To identify causative genes, whole exome sequencing using the Agilent SureSelect 50 Mb Human Exome target enrichment technology was performed in 60 EO PD families confirmed to be consanguineous by their Genomic inbreeding coefficient F values. RESULTS: Linkage analyses using the Genomically Controlled Homozygosity Mapping Statistic under heterogeneity (HLOD) and stratified according to geographical origin have revealed 5 linked regions (7p, 7q, 10p, 1p, 1q) across the genome with HLOD scores ≥2. Using the Illumina cnvPartition module, we did not detect any large rearrangements in the series of 160 consanguineous families. First analyses of the whole exomes of index cases have led to the identification of homozygous mutations in ATP13A2, FBX07. CONCLUSION: As a proof of concept, the strategy combining homozygosity mapping/genomic rearrangement detection and targeted exome sequencing in consanguineous families has led to the identification of homozygous mutations in ATP13A2, FBX07 and probably more novel causative genes for recessive PD.
Rare variant discovery of progressive supranuclear palsy using whole-exome sequencing. C.-F. Lin1-3, E.T. Geller1, O. Valladares1-3, Y.-H. Hwang1,2, L. Stutzbach1, L. Cantwelld, L.-S. Wang1,2, G.D. Schellenberg1,2
1) Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Institute for Biomedical Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Genomics and Computational Biology Graduate Program, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

Progressive supranuclear palsy (PSP) is a rare neurodegenerative disorder that causes movement disability and cognitive decline, and hence is often misdiagnosed as Parkinson’s disease or Alzheimer’s disease. MAPT is a well known genetic factor for PSP with strong effect size. In 2011 a GWAS of 1114 cases and 3247 controls identified SNPs within three genes (STX6, EIF2AK3 and MOBP) with genome-wide significance. However, these common variants do not fully explain the known heritability of PSP.

Whole-exome sequencing (WES) technologies provide a cost-effective approach to genotype more than 44M human exonic regions of the human genome and represents an exciting direction for rare variant discovery. We performed WES of 278 PSP patients using Illumina Hi-Seq 2000 (100nt pair-end) and Nimblegen SeqCap EZ Human Exome library v2 for target capture. Sequencing data were analyzed using the NIAGADS DRAW workflow, following the best practices suggested by the Broad Institute Genomic Analysis Toolkit (GATK) site. All samples have at least 80% of the targeted regions at >20x and 90% of the regions at >10x coverage. Overall we observed 165,768 exonic variants (including core splice sites), 37,731 (23%) of which were not reported in dbSNP 137, -95,000 were not synonymous SNVs. On average each of our patients had 7,1571 (CI:15202-15241) single nucleotide variations (SNVs) and small insertions/deletions. We next examined the variants in a set of 61 genes known to be related to neurodegenerative disorders. We found that a missense SNV in the GAK gene shows significant association: the minor allele is present in five out of 566 alleles (five out of the 278 PSP subjects). Four of the five heterozygotes were confirmed through Sanger sequencing. A fisher exact test with four out of 8598 alleles in the NHLBI GO Exome Sequencing Project (ESP) European American population was significant (p-value of 7.0e-13). A follow-up of a patient with hyperactivity showed elevated number of loss of function variants in the FIG4 gene. Our initial analysis is encouraging as it suggests that rare variants in the exome may contribute to PSP risk. Our next step is to complete whole-exome sequencing of 750 PSP subjects, perform a large-scale genetic analysis and perform gene-wise burden tests using additional controls, and validate called rare variants using independent molecular methods such as Sanger sequencing or TaqMan PCR.

We will report on the replication analyses of these preliminary findings in this presentation.

A recurrent microdeletion at 20p13 unmasks a recessive mutation in PLCB1 in a patient with severe infantile epileptic encephalopathy. I.M. Wentzel1-2, A. Mezaque3, E. Mayer2, C. Appelgeta4, A. Ngoh2, E. Kor2,4, D. Batista2,6, T. Wang1-2, 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Neurosciences Unit, University College of London, Institute of Child Health, London, UK; 3) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Kennedy Krieger Institute, Baltimore, MD.

Phospholipase C beta 1 (PLCB1) is a G protein-coupled phosphodiesterase which catalyzes the formation of second messenger molecules inositol 1,4,5-trisphosphate and diacylglycerol from phosphatidylinositol 4,5-bisphosphate. PLCB1 plays an important role in intracellular signaling pathways in the central nervous system. A recent report described a homozygous 20p13 microdeletion involving the promoter and first three exons of PLCB1 to cause infantile epilepsy in consanguineous families. We report a 10 month-old African American female with intractable epilepsy, severe global developmental delay, and regression. This infant was born at 37 weeks of gestation after a normal pregnancy and delivery. No significant dysmorphic features or congenital malformations were present on physical exam and imaging studies. Metabolic studies were unremarkable. Brain MRI showed mild cerebral volume reduction, mild hypoplastic corpus callosum, age-appropriate myelination, and normal neuronal migration patterns. EEGs showed right temporal lobe sharp waves and spikes followed by generalization of epileptiform activities. A SNP microarray identified a heterozygous intergenic deletion of 476kb (80066081-80072779) spanning some 20p13 involving the promoter region and first three exons of the PLCB1 gene. This deletion was inherited from her mother who has no history of seizures. Long-range PCR was performed to establish the exact breakpoints of the deletion. These are similar to the previously reported first cases of microdeletion at 20p13. Interestingly, these recurrent microdeletions are flanked by two long interspersed (LINE) elements suggesting nonallelic homologous recombination at this location. Sanger sequencing of exons of PLCB1 identified a single base substitution (c.898G>A) which involves the conserved splice site of intron 1 in the proband and her unaffected father. These results support that loss-of-function mutations in PLCB1 cause a rare recessive from of early infantile epileptic encephalopathy and that flanking LINE elements may play a role in recurrent microdeletions in human disease.

A Follow-Up of Whole-Exome Sequencing in Multiplex Families. A.H. Beecham1, J.L. McCauley1, A. Hadjixenofontos1, P.L. Whitehead1, W.F. Hulme1, I. Konidari1, S.L. Hauser2, J.R. Okسنberg2, J.M. Vance1, J.L. Haines3, A. Pericak-Vance1, 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Department of Neurology, School of Medicine, University of California, San Francisco, CA, USA; 3) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN, USA.

Multiple sclerosis (MS) is a common neurodegenerative disease, affecting more than 1.3 million individuals worldwide. Given the number of rare MS families found in MS, it is plausible that rare variants may contribute to MS in these families. We have used next-generation sequencing to scan multiplex families for rare variants. Filtering 127 individuals (91 affected and 36 unaffected) from 26 multiplex families, assuming a dominant model with incomplete penetrance, we identified 578 high quality coding (nonsense, missense, or splice), conserved, potentially damaging, and rare (MAF ≤ 0.01) variants in 550 genes segregating completely in all affected of at least one family. Three of the 578 variants are in confirmed MS genes including CLEC16A, EOMES, and SLC9A8. Assuming a recessive model, we identified 68 high quality coding, conserved, potentially damaging, and rare (0.01 < MAF ≤ 0.05) variants in 61 genes segregating completely in at least one family. Using a gene focused approach; we found 20 genes having at least two different filtered variants segregating in at least two different families. Eight of the 20 genes showed significant burden (p < 0.05 from Fisher’s exact test) in our families when compared to 141 non-diseased controls including: PNLPLA6, PLXNA4, D1H1A, PDH2A, SAM3D, COL11A2, SMAD5, and HIVEP2. Lastly, we identified 510 rare (MAF ≤ 0.01) and 10 moderately rare (0.01 < MAF ≤ 0.05) high quality coding, conserved, potentially damaging variants in known MS genes, including the Major Histocompatibility Complex (chromosome 6 from 26-36 MB) that were present in at least one MS affected individual. We submitted 1,154 unique variants from our three filtering approaches (dominant, recessive, known MS genes) for inclusion in a custom Illumina Infinium HumanExome +BeadChip designed by the International Multiple Sclerosis Genetics Consortium. In total, 1,082 of our nominated variants made it onto the chip after manufacturer quality control, including all priority variants from the eight genes demonstrating significant burden. We are currently genotyping 114 affected individuals from 51 additional multiplex families to screen for these variants and other rare variation in these genes included as part of the exome chip content.
1377T
Exome Sequencing for Schizophrenia and Alcohol Dependence in an Irish Cohort Using Low-cost Library Prep and Target Capture. D. G. Brothawn1, T. Hendon1, T. Bigdeli2, E. Loken1, D. Walsh2, F. A O’Neill1, S. Bacanu1, K. Kendler1, B. T. Webb1, B. Riley1.
1) Virginia Commonwealth University, Richmond, VA; 2) Health Research Board, Dublin, Ireland; 3) Queens University, Belfast, Ireland.

Common and rare variants contribute to alcohol dependence. Identifying specific common alleles and elucidating their function remain difficult due to the small effect sizes. Rare functional variation in relevant genes may have more direct impact on trait risk and may elucidate the role of variation in the genetic pathogenesis of schizophrenia. In this study, we attempted to detect rare variants associated with schizophrenia and alcohol dependence.

1378F
Reduced mRNA expression and aberrant intron DNA methylation of EGR2 in female schizophrenia patients. M. Cheng1, Y. Chiang1, S. Hsu1, C. Chen2.
1) Department of Psychiatry, Yuli Veterans Hospital, Hualien, Taiwan; 2) Department of Psychiatry, Chang Gung Memorial Hospital at Linkou and Chang Gung University School of Medicine, Taoyuan, Taiwan.

Background: Abnormal myelination has been considered as part of the pathophysiology of schizophrenia. Early growth response 2 (EGR2) has a specific function in the regulation of hindbrain development, myelination of peripheral nervous system, and the stabilization of long-term potentiation. And it has been considered as a candidate gene for schizophrenia. Methods: We compared the EGR2 mRNA levels in lymphoblastoid cell lines between 119 schizophrenic patients and 114 controls using real-time quantitative PCR. We attempted to detect genetic variants of the EGR2 gene in 500 patients with schizophrenia and 500 non-psychotic controls from Taiwan using direct sequencing and conducted a case-control association study. We measured the EGR2 mRNA in SH-SY5Y cells treated with a DNA methyltransferase inhibitor, 5-azacytidine, for 24 hours. We also compared DNA methylation status of EGR2 gene in the lymphocytes in between 61 patients with schizophrenia and 60 controls using Reduced representation bisulfite sequencing (RRBS) protocol for unique reads (87.2% baseline vs. 86.1% pooled) and % of bases on or near bait (89.2% baseline vs 91.5% pooled). We have completed prep for 73 additional samples and we anticipate that >100 exomes will be complete and analyzed by the time of presentation.

1379W
Association of rare variants with anorexia nervosa by whole exome sequencing. J. Connolly1, D. Li1, H. Hakonarson1,2.
1) Ctr Applied Genomics, Children’s Hosp Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

Anorexia nervosa (AN) is a perplexing illness characterized by low body weight and persistent fear of weight gain during period of growth, resulting in eating disconnection. People with AN usually manifest with symptoms of depression, anxiety, and obsessive-compulsive behaviors that are common features in other neuropsychiatric disorders. Multiple bodies of evidence now suggest the role of genetic influences to AN. Family studies have consistently demonstrated that AN occurs in families and twin studies have revealed the contribution of additive genetic factors to the observed familial aggregation, but despite many studies conducted so far, the genetic architecture underlying AN susceptibility remains largely unknown. Here we have completed analysis of 10 extreme cases samples and multiple control samples with where an average coverage of 65X was established. A total of 133,955 single nucleotide variants (SNVs) and 70,678 small insertions/deletions (indels) were identified in the 10 case samples that underwent whole exome sequencing (WES), of which about 4% of the variants are novel. Two novel variants were identified in previous candidate genes that involved in the mood control and two in previous genome-wide association signals. Also, we performed association analysis using both internal and 1000 genomes data as control after combining several Caucasian populations. Due to excessive rare variants in WES results, burden test was applied. The total number of rare variants across a gene is tabulated in each individual and these totals are compared between cases and controls. The top association signals were detected in C7orf23, DMTF1, COL18A1, PCNT and ERMN. Interestingly, two PCNT gene variants were previously associated with an increased risk of developing psychiatric disorders such as schizophrenia, and depression was found to be significant (3.65E-03).
1381F Genome-wide profiling of DNA-protein interactions and multiple histone methylation in the cultured olfactory cells from Schizophrenia. Z. Dong1, M. Deng1, O.V. Evgrafov1, J.A. Knowles1. 1) Psychiatry & Behavioral Sci, Univ Southern California, Los Angeles, CA; 2) Undergraduate, Brown University, Providence, RI.

Schizophrenia (SCZ) is a seriously heritable complex disorder. The discovery of candidate risk genes and genetic risk factors for schizophrenia has proven challenging. The great difficulty in the search of genes and biomarkers for schizophrenia can be ascribed to a combination of high disease heterogeneity, the absence of prominent and unique alteration, and the complex interplay of genetic predisposition and environmental influences. We hypothesize that epigenetic variations (DNA methylation and histone modifications), together with particular genetic variations (SNPs, insertion/deletions), play roles in the etiology of Schizophrenia. Furthermore, epigenetic modifications are responsive to environmental factors (including anti-psychotic treatments) and could be a link between genetic predisposition and environmental influence on the development of schizophrenia. To reduce the complexity and limit genetic heterogeneity, we choose a subgroup of SCZ with olfactory deficits and focused on the epigenetic alteration using a genome-wide ChIP-seq of the neuronal precursor cell culture from SCZ nasal biopsy samples. We have performed seven different genome-wide epigenetic profiles (H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K9ac, CTCF and RNA pol II) in four SCZ patients and two controls. We observed more epigenetic heterogeneity among the SCZ samples, while the profiles were more consistent within the control samples. We find several abnormal epigenetic peaks (genome-wide FDR<0.01) are associated with some new schizophrenia loci identified in a mega-analysis combining genome-wide association study (GWAS) data from 17 separate studies. (Schizophrenia Psychiatric Genom-Wide Association Study Consortium, 2011). These epigenetic profiles have been subsequently analyzed together with genome-wide transcriptional profiles from the RNA-seq results of the cases and controls. We have compiled a list of candidate genes, biomarkers and chromosomal regions identified in our epigenetic study and transcriptional study.

1382W Examining De-novo mutation rates and patterns from whole-exome sequencing of specific cases of Schizophrenia from Taiwan. D. Howrigan1,2, B. Neale1,2, J. Moran1, K. Chambert1, S. Rose2, N. Laird3, H.-G. Hwu4, W.-J. Chen5, C.-M. Liu3, J. Nemesh2, E. Bevilacqua2, A. Hansen1, S. V. Faraoe6, S. Glat7, M. Tsuang1, M. McCrannel2, 1) Massachusetts General Hospital, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) SUNY Upstate Medical University, Syracuse, NY; 4) University of California, San Diego, CA; 5) National Taiwan University, Taiwan; 6) Harvard School of Public Health, Boston, MA; 7) Harvard Medical School, Boston, MA.

Recent studies have implicated that exonic de-novo mutations play a convincing role in developmental disorders such as autism, intellectual disability, and epilepsy. For sporadic cases of schizophrenia, early reports have pointed towards promising de novo candidates: however larger numbers of sequenced trios are required to identify specific genes, as de novo events arise at a low rate in the population (roughly one per trio). Under the collaboration of multiple centers, whole-exome sequencing has been performed on 1,135 complete trios from a Taiwanese cohort, making it one of the largest sequenced trios are required to identify specific genes, as de novo events (those with twenty or more reads in the full trio), multiple loss-of-function mutations found in SV2B, a synaptic vesicle glycoprotein, were identified. We studied multiple loss-of-function mutations found in SV2B, a synaptic vesicle glycoprotein gene primarily expressed in the brain. These findings, however, do not surpass exome-wide significance after incorporating gene size and site-specific mutation rates into expectations of de novo mutation. Further analyses were done on a large dataset of 52 cases with schizophrenia using the Agilent SureSelect exome capture platform, and the Illumina MiSeq platform was chosen as the method for validating putative de novo calls. Preliminary results suggest that the overall rate of mutation in affected offspring falls in line with the expected mutation rate. Upon examination of high-quality putative de novo events (those with twenty or more reads in the full trio), multiple de-novo mutations have been found in several genes. We observed higher enrichment of loss-of-function events and a 1.19-fold higher enrichment of non-synonymous events. However, these levels only reach suggestive significance (p = 0.08 and 0.1, respectively). Overall, our findings do not clear the gene against an uninformative risk factor for schizophrenia when disrupted by de novo mutation, although a number of genes and gene-set analyses show suggestive signals. We also integrate the results from these trios with the other available trio sequencing data from the consortium, to provide a clear path of the role of the de novo events play in the etiology of schizophrenia.


Recent studies have found heterozygous rare variants in TREM2 (triggering receptor expressed on myeloid cells 2), which has been reported to be involved in inflammation in the brain, significantly affect risk for late-onset Alzheimer’s disease (LOAD). Although several rare variants have been associated with LOAD risk, large scale deep re-sequencing studies have not been conducted to comprehensively identify novel rare variants in TREM2 that affect LOAD. We performed pooled-DNA sequencing of all exons and flanking regions in TREM2 in 2,074 LOAD cases and 1,388 cognitively normal elderly controls of European descent from the National Cell Repository for Alzheimer’s disease, the Knight Alzheimer’s Disease Research Center, and the Alzheimer’s Disease Neuroimaging Initiative using Illumina Miseq sequencing. We used the SPLINTER algorithm to perform alignment and call variants in the pooled samples. SIFT2, SeattleSeq and Ensembl databases were used to annotate variants. We then confirmed rare variant calls in all sequenced individuals using Sequenom iPLEX or KASPar genotyping systems. Of the 122.8 million reads generated, 69.7 million (56.7%) mapped back to the reference genome with a 30-fold minimum coverage. Eight novel or potentially functional rare variants passed the filtering quality controls and were selected for direct genotyping. Six of the eight variants were validated (sensitivity=75%), four of which are novel. Five rare variants (4 missense and 1 stop-gained) were present only in cases and not in controls. Among them, four were predicted as probably-damaging by PolyPhen or SIFT2. We also observed one missense mutation present only in controls and predicted to be tolerated. When collapsed, these six rare variants in TREM2 were associated with a significant increase in LOAD risk (OR=6.04; P=0.04) which suggests that some could be functional and have dramatic effects on LOAD risk. This study suggests deep re-sequencing is an effective and accurate approach to identify novel rare variants associated with LOAD. The identified heterozygous rare variants may cause a loss-of-function of TREM2, consistent with the notion that heterozygous TREM2 mutations contribute to LOAD.

1384F Mutations in the CASR gene in Idiopathic Generalized Epilepsy. M. Kaur1, P. Satishchandra2, K. Radhakrishnan3, A. Kapoor4, A. Anand3, 1) Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, Karnataka, India; 2) Department of Neurology, National Institute of Mental Health and Neurosciences, Bangalore, Karnataka, India; 3) Department of Neurology, Sri Chitra Tirunal Institute for Medical Sciences and Technology, Trivandum, India.

Idiopathic generalized epilepsy (IGE) is a common type of epilepsy with substantial genetic basis to its etiology. Genetic studies have identified 12 IGE loci, and at 6 of these, causative genes have been found. We conducted a detailed sequence analysis of an IGE locus, EIF8 (MIM612999) which was identified in a family from south India with several of its members affected with epilepsy (Kapoor et al. Ann Neurol 2008). EIF8 maps at chromosome 3q13-q21. We used a combination of Sanger sequencing and next generation sequencing to analyze transcripts for 234 genes in the critical genomic region. New single nucleotide variants were found in the NSUN3, EPHA6, ABIP3BP, TRAT1, KIAA1407, IQCB1, CASR and ADCY5 genes. We studied segregation of the variants in the epilepsy family, examined their allele frequencies in the control individuals and conducted in silico pathogenicity analysis. There is one missense variant in CASR (extracellular calcium sensing receptor) in causation of epilepsy in the family. Additionally, we examined CASR in 480 IGE/JME patients and 252 control individuals, and detected 22 uncommon variants (m.a.f≤0.005). Of the variants identified, 7 were missense variants leading to E354A, D433H, S580N, I686V, R898Q, A988S, A988V. While CASR is known to play a role in maintenance of systemic calcium homeostasis, its role in neuronal cells remains unclear. CASR may play a role in regulating neuronal excitability in the human brain. We studied functional correlates of the missense variants identified using a C. elegans path model and an iP-Orchestrated ELISA-based assay in HEK293T cells, transiently transfected with the wild type and mutant receptor constructs. Three mutant receptors exhibited significantly enhanced Ca2+ responsiveness as compared to wild type CASR, whereas the remaining ones, showed relatively weak significant differences in Ca2+ responsiveness as compared to wild type CASR, where the remaining one, showed relatively weak significant differences in Ca2+ responsiveness as compared to wild type CASR, which may be attributed to the variants.

Posters: Psychiatric Genetics, Neurogenetics and Neurodegeneration 299
1385W

Tri-o based pathway analysis of bipolar disorder. N. Matoba1,2, M. Kataoka1,2, K. Fujiy4, Y. Suzuki2, S. Sugano2, T. Kato1. 1) Lab. for Molecular Dynamics of Mental Disorder, RIKEN Brain Science Institute, Wako, Saitama, Japan; 2) Dept. of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan; 3) Dept. of Child Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Hong-ko, Tokyo, Japan; 4) Dept. of Psychiatry, Dokko Medical University, Mibu, Tochigi, Japan.

Bipolar disorder is one of the two major mental disorders and twin studies reported the heritability is around 85% (Cardno et al., 1999; McGuffin et al., 2003). Genome-Wide Association Studies (GWAS) identified a number of common SNPs that are associated with bipolar disorder, but no strong contribution has been shown. Next-Generation Sequencing (NGS) technology enables us to detect causes of rare genetic disorders. However, the way to identify causative genes of complex diseases like bipolar disorder is still under development. To identify molecular pathways related to bipolar disorder, we compared the deleterious mutations transmitted to, or un-transmitted to the proband using whole exome sequencing of trio families of bipolar disorder. Fifty trio families were enrolled. Each individuals was interviewed by trained psychiatrists using a structured interview were selected. After target exome were captured from the whole-blood or saliva DNA of participants using the SureSelect Human All Exon V4(Aglient), whole-exome sequencing was carried out using HiSeq2000(Illumina). Reads were mapped to human genome build v37 with BWA and PCR duplicates were removed with Picards. Local re-alignment and variant calling was performed using GATK. All variants were annotated with ANNOVAR. 87.8% of target regions were covered with at least 20 unique reads and we only focused on SNPs with a minor allele frequency (MAF) < 1.01 as disease related and predicted as ‘disease causing’ by SIFT and ‘possibly damaging or probably damaging’ by PolyPhen2. All variants observed in each family were divided into two groups, transmitted (found in a proband and one of the parents) or un-transmitted (found in a parent only) variants. In a preliminary analysis of 7 families, 42.86 variants were transmitted to proband and 49.29 variants were un-transmitted to the proband on average. Gene Ontology (GO) terms related to GO terms were enriched in transmitted variants while muscle development related GO terms were enriched in un-transmitted variants. Enrichment of apoptosis related genes in transmitted mutations is compatible with the possible role of cellular viability in bipolar disorder. Further analysis using other 43 trio families with bipolar disorder and control trio families is ongoing.

1388W


Autism Spectrum Disorders (ASD) are a group of developmental disorders that involve three specific developmental areas: disrupted communication, difficulties in social interaction and repetitive/stereotyped behavior. Even though its etiology is still unknown, evidence suggests a strong genetic component evidenced by twin studies that demonstrate approximately 80% concordance. However, only about 10-20% of non-syndromic ASD cases have been explained by deleterious mutations, but each genetic effect explains only up to 1% of the ASD cases. Contactin Associated Protein-Like 2 (CNTNAP2) [MM: 604589], encodes a protein from the Neurexin family with various roles in the Central Nervous System. The aim of this study was to search for deleterious variations (no-synonymous, low frequency, rare or private) in the CNTNAP2 gene in ten individuals with ASD by sequencing the coding region of this gene using Sanger technique. We found 25 variations in the CNTNAP2 gene, but we focused on a particular haplotype (rs6123377, rs7702588, and rs9648691) that was found in 4 individuals containing the intron 22/exon 23 region. According to bioinformatics analysis this haplotype suggests a formation of a splicing branch site on exon 23. This could be possibly involved in the formation of new isoforms of the gene, and therefore might be related to ASD.

1388T

The not-so-silent effect of silence mutations in Autism Spectrum Disorders. O.A. Moreno-Ramos, M.C. Lattig, Liga Colombiana de Autismo. Universidad de los Andes, Laboratorio de Genetica Humana, Bogota, Colombia.

High locus and allelic heterogeneity found in Autism Spectrum Disorders (ASD) complicate the comprehension of the genetic bases of the disorders. Thousands of clues have emerged from initial exome sequencing studies, identifying many novel de novo mutations in ASD family trio studies. Affected proteins seem to be highly interconnected and expressed mainly in brain. Most of the variants reported to date were found mostly in Caucasian or European descendant cohorts but Latin American cohorts have not been well studied. Therefore, we decided to apply exome sequencing, at a 50X depth, in a cohort of Colombian - South American (admixed population) trios. Although most of exome sequencing studies focus on deleterious mutations such as non-synonymous, missense and frame shift mutations, there is a still a big caveat: What about the not-so-silent effect of silence mutations might have? We not only focused on the discovery of harmful variants, but we also evaluated the possible effect of silent mutations, since it is known that synonymous mutations can actually be a cause of different diseases and syndromes altering mRNA stability or translation. The global outcomes guide to a larger range of mutations that are related to ASD.

1387F


Recently, two independent research groups (Jonsson et al. 2013 and Guerreiro et al. 2013) published results reporting a rare missense variant rs75932628 (R47H) in exon 2 of the gene encoding triggering receptor expressed on myeloid cells 2 (TREM2) significantly increase the risk for Alzheimer’s disease (AD) with an effect size comparable to that of the APOE e4 allele. Subsequently, Pottier et al. (2013) and Benitez et al. (2013) published independent short communications confirming the association between the minor T-allele at rs75932628 and increased risk for AD. Here we attempt to replicate the association between rs75932628 and AD risk by directly genotyping rs75932628 in two independent Caucasian family cohorts, the National Institute of Mental Health Alzheimer Disease Genetics Initiative Family Study and the National Institute on Aging Genetics Initiative for Late Onset Alzheimer’s Disease (NIA-LOAD) Family Study, consisting of 927 families (with 1777 affecteds and 1258 unaffecteds) and in the Caucasian NIA-LOAD case-control sample composed of 378 cases and 686 controls. Additionally, we imputed genotypes in three independent Caucasian case-control cohorts (GenADA, TGEN2, and ADNI), containing 1906 cases and 1503 controls. Meta-analysis of the two family-based and the four case-control cohorts yielded a P-value of 0.04 providing additional independent support for the association between the T-allele at rs75932628 and increased AD risk, albeit suggesting a much lower effect size (OR = 1.48, 95% CI 1.02-2.13). Our results confirm the association between SNP rs75932628 and AD risk, however, the risk effect is substantially smaller than in the two original reports.
1389T


1) Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 2) BGI-Shenzhen, Shenzhen, China; 3) Neurology Service and Stroke Unit, General Hospital S. Michele AOB “G. Brotzu”, Cagliari, Italy; 4) Neurology Division, S. Francesco Hospital, ASL No. 3 Nuoro, Italy; 5) Clinical Pharmacology Unit, Instituto de Medicina Molecular, University of Lisbon, Portugal; 6) Parkinson Institute, Istituti Clinici di Perfezionamento, Milan, Italy; 7) Neurology Service, Hospital Clinic of Barcelona, Barcelona, Spain; 8) Institute of Neurological Sciences, National Research Council, Cosenza, Italy; 9) Department of Medical Biotechnologies and Translational Medicine, University of Milan, Milan, Italy; 10) Institute of Neurology, University Magna Gracia, Catanzaro, Italy; 11) Neuroimaging Research Unit, Institute of Neurological Sciences, National Research Council, Germaineto, Italy; 12) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 13) King Abdulaziz University, Jeddah, Saudi Arabia; 14) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark.

Parkinson’s disease (PD) is the second most common neurodegenerative disorder after Alzheimer’s disease, with a prevalence of >1% over age 65. The etiologic landscape of PD is complex. Rare, highly-penetrant mutations in different genes and common risk factors of small size-effects in several loci have been identified in patients with PD. However, additional disease determinants remain to be identified in most patients. Genetically isolated populations offer advantages for dissecting the genetic architecture of complex disorders. Here we report the results of our genetic study on the etiology of PD in Sardinia. We performed exome sequencing in 100 unrelated Sardinian PD patients, using the Agilent SureSelect 50Mb kit and Illumina HiSeq2000 sequencing, at an average coverage of ~30x. We first removed all synonymous, intronic, and intergenic variants. From the remaining variable sites, we then identified variants present only in PD patients, and absent in dbSNP129 and 1000Genomes databases. This approach yielded a total of 4,587 SNPs, that were then genotyped in 500 independent Sardinian individuals (242 PD and 258 controls) using a NimbleGen SeqCap EZ custom platform. Association of each variant with disease status was then tested using Fisher’s exact test implemented in PLINK/SEQ. Out of 186 variants with p-value <0.5 and odd-ratio >3, thirty-two were confirmed by Sanger sequencing. These 32 SNPs were then genotyped by TaqMan assays in an independent sample of 2,731 PD patients and 2,673 age- and ethnic-matched controls from Italy, Spain and Portugal. No variants surpassed the required level of significance according to Bonferroni correction for 4,587 tests (p value < 1.09 x 10^-6). The lowest p value was 1.16 x 10^-6 for a variant with final OR 2.9. However, a catalogue of interesting variants identified were further screened in 260 controls. [Results] Three of 11 FHM patients (27.3%) harbored causative mutations, including a previously reported mutation p.T666M in CACNA1A in 2 patients and a novel mutation p.H916L in ATP1A2 in a patient. In addition, a known causative mutation c.649dupc in PRRT2 was identified in a SHM patient. No mutations were found in patients with MA, MO or EA. Common clinical features of FHM patients with the mutations included relatively long-duration of attacks, disturbance of consciousness, fever, and cerebral edema during hemiplegic attacks. [Discussion and Conclusion] This study indicated that a substantial number of patients with hemiplegic migraine in the Japanese population were accounted for by mutations in known causative genes including CACNA1A, ATP1A2 and PRRT2. In addition, this study have raised the possibility that p.T666M in CACNA1A might be a relatively frequent mutation in FHM in the Japanese population, as supported by the finding that the two of three previous reports with mutations in CACNA1A in the Japanese population represented the p.T666M mutation. Genetic epidemiological studies in other populations may be needed to confirm this finding. [Conclusion] New disease genes can be identified through high throughput sequencing approaches and these will likely be present in populations including those with rare or undiagnosed cases. Further studies are needed to examine the potential for candidate prioritization in PD and other complex diseases.

1391W

Whole transcriptome analysis by next generation sequencing in autism spectrum disorders. C. Zusi1, P. Prandini1, G. Malerba1, L. Xumerle1, R. Galavotti1, A. Pasquali1, C. Patuzzo1, V. Mijatovic1, R. Ciccone2, M. Ficher3, MC. Bonaglia1, E. Trabetti1, PF. Pignatti1, 1) University of Verona, Verona, Verona, Italy; 2) University of Pavia, Pavia, Italy; 3) Scientific Institute Oasi Maria Santissima, Troina, Italy; 4) Scientific Institute E. Medea, Bosisio Parini, Italy.

Autism Spectrum Disorders (ASDs) represent a group of childhood neuro-developmental and neuropsychiatric disorders characterized by deficits in verbal communication, impairment of social interaction, and restricted and repetitive patterns of interests and behaviours. Evidences indicate that ASDs have strong genetic bases. This study is part of a Telethon project involving several Italian research groups; it aims to analyze differences of gene expression level between 27 ASD subjects and 23 control individuals. Nineteen (19) of the 27 ASD subjects are characterized for CNVs potentially involved in the onset of autism and 8 subjects have a deletion in the 22q13.3qter region. EBV transformed lymphoblastoid cells have been established for all subjects and transcriptomes have been analyzed through Next Generation Sequencing technology (RNA Sequencing). Differential expression analysis, performed on the whole cohort and the subgroup with the 22qter deletion, identified 295 and 448 differentially expressed genes respectively, at the nominal p-value <0.05. Gene set enrichment analysis (GSEA) revealed that autoimmune disorders and antigen processing and presentation terms are the most enriched ones. Subgroup’s GSEA highlights the involvement of axon guidance term, confirming that LLCs could exhibit brain markers relevant to autism. Further analysis was carried out to identify genes strongly dysregulated in ASD subjects, finding a total of 68 genes. Three dysregulated genes, that cluster within a CNV on chromosome 16p13.1, were identified in one subject. The remaining 65 dysregulated genes did not map on any considered CNV. Additionally, the differentially expressed genes will help understanding the genetic bases of ASD pathophysiology and unravelling potential new pathways involved in the disease.

Purpose: Autosomal-dominant optic atrophy (DOA) is one of the most common inherited optic neuropathies causing progressive bilateral degeneration of the optic nerve and ultimately irreversible blindness. Mutations in OPA1, encoding a dynamin-like mitochondrial GTPase indispensable for mitochondrial network structure and morphology, are identified to be responsible for DOA in approximately 50% of cases. Genetic abnormalities responsible for the remaining 50% of cases are not yet known. OPA1 antisense RNA 1 (OPA1-AS1) is a long non-coding mRNA located between intron 5 and 7 of the OPA1 gene that may function to regulate the expression of OPA1. To investigate the association between OPA1-AS1 variants and DOA, we performed a sequencing-based association study. Methods: The study was approved by the Massachusetts Eye and Ear Infirmary Institutional Review Board and adheres to the tenets set forth in the Declaration of Helsinki. Fifty-nine DOA patients from 53 independent families were sequenced in this study. Disease causing OPA1 mutations were previously detected in 17 of these 59 patients. Genomic DNA was sequenced using primers designed to amplify all 3 OPA1-AS1 exons. PCR products were sequenced according to standard protocols. Genotypes of known SNPs in the general population were retrieved from the HapMap CEU data sets (n = 174). Results: Four OPA1-AS1 SNPs were observed in cases: rs9291059, rs34307082, rs3772393, and rs9832709. Of the 59 DOA patients, 2 patients were heterozygous for rs9291059 [minor allele frequency (MAF) = 0.018] and one had a homozygous deletion for rs34307082 (Mut). Significant linkage disequilibrium was observed between common SNPs rs3772393 and rs9832709 (D = 1.0 and r^2 = 1.0). We found that the MAF of OPA1-AS1 rs3772393 was significantly higher in the DOA cases than that in general population controls (0.53 versus 0.42, P = 0.004, odds ratio = 1.77, 95% confidence interval = 0.84-3.76). Of the 16 patients homozygous for the rare variant of rs3772393, 7 (43.8%) have a disease-causing OPA1 mutation, although none of these were located in the OPA1-AS1 exons and are not in linkage disequilibrium with rs3772393. Novel variants in OPA1-AS1 were not identified by sequencing. Conclusions: These data suggest that OPA1-AS1 variants may associate with developing DOA and that additional study of the potential role of OPA1-AS1 in DOA is warranted.

1393F Exome Sequencing on Samples with Bipolar Disorder: A Preliminary Survey. T. Zhang1, F.J. McMahon2, D.T. Chen2, J.C. Wang1, J.P. Rice1. 1) Dept. of Psychiatry, Washington University in St. Louis, St. Louis, MO; 2) Genetic Basis of Mood and Anxiety Disorders Unit, National Institute of Mental Health, Rockville, MD.

Bipolar disorder is a psychiatric diagnosis for a mood disorder with lifetime prevalence of about 1%. Previous genetic studies have linked many chromosomal regions and candidate genes appearing to influence the risk of bipolar disorder. In this research, we aim to identify the potential causal variants for bipolar disorder via performing exome sequencing on 6 related case samples from a four-generation family. The sequencing data were released for 195,949 variants, and 20,094 variants (10.25%) were shared among all these six related cases. Genotype data annotation was performed by Annovar and by implementing several standard filtering criteria we have reduced the number of targeted variants to 9,657 variants. To narrow the susceptible regions, we have genotyped 735,202 common SNPs for all these six cases and performed linkage analysis. A preliminary low-density linkage analysis has identified three suggestive peak region in chromosomes 4, 8 and 10 (with LOD score greater than 2.0). We are now performing high-density/haplotype-based linkage analysis to further narrow down the candidate regions and identify the potential susceptible variants for bipolar disorder in this family.

1394W Exploration of brain-specific somatic mutations with massively parallel sequencing data derived from human postmortem brain. M. Nishioka1,2, M. Bandou1, J. Ueda2, S. Murayama4, K. Kasaia2, T. Kato4, K. Iwamoto1. 1) Department of Molecular Psychiatry, The University of Tokyo, Tokyo, Japan; 2) Department of Neuropsychiatry, The University of Tokyo, Tokyo, Japan; 3) Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, Saitama, Japan; 4) Department of Neuropathology, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan.

Accumulating evidence has challenged the conventional assumption that somatic cells in several individual, except for rare cases, have the identical genome. Several groups have reported somatic single nucleotide variations (SNVs) and structural variations (SVs) in normal cells other than oncogenic cells in humans and model animals. Especially, brain cells are supposed to go through characteristic genomic alterations such as retrotransposition and chromosomal aneuploidy during development. We hypothesized that such somatic genomic alterations in brain cells contribute to biological diversity in cognitive and mental traits, and that some somatic mutations are involved in the pathophysiology of neuropsychiatric disorders. We explored brain-specific somatic SNVs and SVs with exome sequencing data derived from human postmortem brain and liver with no neuropsychiatric disorders as a preliminary study. The samples are fresh-frozen postmortem frontal cortex and liver derived from 75-year-old Japanese male. We obtained neuronal genomic DNA by anti-NeuN antibody-based FACS technique. The exome sequencing of neuronal and liver genomic DNA was performed with Illumina GAIIx. The total reads were approximately 100 million reads per each sample, with more than x100 coverage. We analyzed somatic mutation events using the generated read data with SNV/SV detection software such as RetroSeq, Delly, and SVDetect. We found that sequencing data derived from neuronal nuclei indicated much higher rate of retrotransposition and LTR-related genomic alterations. Although such genomic alterations might be results of sequencing or informatics error, some would be results of somatic mutations occurring very low fraction of sample tissues. We discuss the physiological and pathophysiological meanings of brain-specific somatic mutations and the validity of detection methods, mainly focusing on the results of transposable elements.
1395T

Since discovery of the SMN gene in 1992, many rare clinical variants of SMA have been reported, including a subset associated with AMC. The genetic basis of these disorders remains unknown and continues to present important diagnostic challenges in the newborn period. In 2008, we uncovered the genetic cause of one of these rare, infantile forms (X-linked SMA (SMAX2, OMIM 301830)) through identification of the first human mutations in UBA1. We continue to evaluate families and affected male cases (AM) with XL-SMA-like phenotypes with the goal of discovering the genetic bases of these diseases. For the past year, our approach has been targeted sequencing of UBA1, followed by whole exome sequencing of UBA1-mutation negative cases. We have sequenced all UBA1 coding regions in all affected individuals in our cohort using a custom designed assay on Ion Torrent. As expected, UBA1 sequencing revealed common variants in each individual, but surprisingly no new disease associated variants were identified. We performed whole exome sequencing of 37 samples (15 families including 15 AM) on Illumina HiSeq 2000/2500. Alignment and extensive variant analysis was performed using custom in-house pipelines. Selected variants of interest were validated by qPCR Taqman assays. Exome sequencing revealed probable pathogenic mutations in many cases; we highlight results from three of these. First, a family with two affected boys revealed novel compound heterozygous deleterious CHRNA7 mutations (OMIM 100720) that cause a form of lethal congenital myasthenic syndrome. Second, a novel start loss M1V mutation in SCML2 (OMIM 300208) was detected in the proband and mother of a family with an X-linked history of fetal and neonatal deaths. SCML2 encodes a component of the polycomb transcriptional repressor complex, and has not been previously associated with human any disease. Third, in a singleton affected boy, we identified a translocation of exons 21 and 22 of ATRX (OMIM 138100), which results in a frameshift in exon 22 and has not been previously reported in this context. We validated that these mutations co-segregate with disease in this family.

Exome sequencing revealed probable pathogenic mutations in many cases; we highlight results from three of these. First, a family with two affected boys revealed novel compound heterozygous deleterious CHRNA7 mutations (OMIM 100720) that cause a form of lethal congenital myasthenic syndrome. Second, a novel start loss M1V mutation in SCML2 (OMIM 300208) was detected in the proband and mother of a family with an X-linked history of fetal and neonatal deaths. SCML2 encodes a component of the polycomb transcriptional repressor complex, and has not been previously associated with human any disease. Third, in a singleton affected boy, we identified a translocation of exons 21 and 22 of ATRX (OMIM 138100), which results in a frameshift in exon 22 and has not been previously reported in this context. We validated that these mutations co-segregate with disease in this family.

1396F
Genetics of spinocerebellar ataxias in Portuguese families: screening for SCA15, SCA28 and SCA36. J.R. Loureiro1, A.I. Seixas2, J.L. Loureiro1, A. Carracedo3, M.J. Sobrito4, P. Coutinho1, J. Sequeiros5, I. Silveira5, 1) IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Portugal; 2) Serviço de Neurologia, Centro Hospitalar entre Douro e Vouga, Portugal; 3) Fundación Pública Galicia de Medicina Xenómica-SERGAS, Santiago de Compostela, Spain; 4) Genome Medicine Group, School of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain; 5) ICBAS, Universidade do Porto, Portugal.

The spinocerebellar ataxias are rare neurodegenerative diseases clinically and genetically very heterogeneous. To this moment, 32 autosomal dominant SCAs have been genetically identified and 24 causative genes implicated in these pathologies. Trinucleotide repeat expansions are the cause of nine SCAs (SCA1, 2, 3, 6, 7, 8, 12, 17 and DRPLA). Intronic pentanucleotide repeats, an ATCTC expanded tract or an insertion, are also causative of these pathologies (SCA10 and 31). Besides repeats, classical mutations are the origin of at least 12 of these diseases (SCAS 11, 13, 14, 15, 18, 19/22, 23, 26, 27, 28, 35). More recently, intronic GGCGTG hexanucleotide expansions have been found in SCA26. Mutation screening in known SCA genes allowed us characterize approximately 200 Portuguese families. Machado-Joseph disease/SCA3 is the most frequent (52%) followed by DRPLA (5%) and SCA2 (2%), whereas the remaining showed very low frequencies or were not found. Thus, more than 100 Portuguese SCA families remain without a molecular diagnosis. Here we present the results of mutation screenings for SCA15, SCA28 and the recently described SCA36 in these Portuguese families. All previously reported SCA15 patients have deletions of several exons of the ITPR1 gene; SCA28 is caused by missense mutations in or small indels in the AFG3L2 gene; GGCGTG hexanucleotide expansions, ranging from 650-2500 repeats, in intron 1 of NOP56, cause SCA36, the most common type of SCA in Galicia, the Spanish region on the border with Northern Portugal. We carried out quantitative real-time PCR, direct sequencing of exons of interest or repeated-primed PCR to screen for ITPR1 genomic deletions, AFG3L2 mutations and NOP56 expansions, respectively, and failed to find pathogenic alterations. In conclusion, SCA15, SCA28 and SCA36 are very rare among Portuguese SCA families and most probably a considerable number of SCA genotypes remain to be identified. Funding: FCT (Project PTDC/SAU-GMG/098305/2008) and COMPETE with co-funding by FEDER; Instituto de Salud Carlos Ill-FIS PI12/00742.

1397W
Repeat interruptions in spinocerebellar ataxia type 10 expansions are strongly associated with epileptic seizures. K.N. McFarland1, J. Liu1, I. Landriani, S. Raskin1, M. Mosovich1,2, E.M. Gatto3,5, H.A. Teive5, A. Ochoa6, A. Rasmussen3, T. Ashtizawa4, 1) Dept of Neurology and The McNichot Brain Institute, University of Florida, Gainesville, FL; 2) Core for Advanced Molecular Investigation, Graduate Program in Health Sciences, Center for Biological and Health Sciences, Pontifical Catholic University of Paraná, Curitiba, PR, Brazil; 3) Movement Disorder Unit, Neurology Service, Hospital de Clinicas, Federal University of Paraná, Curitiba, PR 80060-150, Brazil; 4) Departamento de Neurologia, Sanatorio de la Trinidad Mitre, Buenos Aires, Argentina; 5) Instituto de Neurociencias Buenos Aires, INEBA, 1429 Buenos Aires, Argentina; 6) Department of Neurogenetics, Instituto Nacional de Neurología y Neurocirugía Manuel Velasco Suarez, Mexico City, DF, Mexico; 7) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK.

Spinocerebellar ataxia type 10 (SCA10), an autosomal dominant neurodegenerative disorder, is the result of a non-coding, pentanucleotide repeat expansion within intron 9 of the Ataxin 10 gene. SCA10 patients present with pure cerebellar ataxia; yet, some families also have a high incidence of epilepsy. SCA10 expansions containing pentanucleotide interruption motifs, termed ATCTC interruptions, experience large contractions during germline transmission, particularly in paternal lineages. At the same time, these alleles confer an earlier age at onset which contradicts traditional rules of genetic anticipation in repeat expansions. Previously, ATCTC interruptions have been associated with a higher prevalence of epileptic seizures in one Mexican-American SCA10 family. In a large cohort of SCA10 families, we analyzed whether ATCTC interruptions confers a greater risk for developing seizures in these families. Notably, we find that the presence of repeat interruptions within the SCA10 expansion confers a 6.80-fold increase in the risk of an SCA10 patient developing epilepsy and a 12.92-fold increase in having a positive family history of epilepsy. We conclude that the presence of repeat interruptions in SCA10 repeat expansion intervals is a significant risk for the epilepsy phenotype and should be considered during genetic counseling.
Androgen receptor gene polymorphism is associated with impulsivity in women with alcoholism. A.M. Manzardo, D. Mettman, E.C. Penick, A.B. Poje, M.G. Butler, Psychiatry and Behavioral Sciences, Kansas University Medical Center, Kansas City, KS.

The androgen receptor (AR) gene, located on the X chromosome, contains a common polymorphism involving CAG repeats in exon 1 which impacts disease and could contribute to the unequal gender ratio in alcoholism. CAG repeats in the AR gene influence androgen sensitivity and are known to correlate with impulsivity in males. CAG repeat lengths over 35 produce a progressive neurodegenerative disorder (Kennedy disease) in men. We report the first preliminary study examining the association between the number of CAG repeats and measures of impulsivity in females with chronic alcoholism. Thirty-five women and 85 men with chronic alcoholism were previously recruited for a nutritional clinical trial. Twenty-six well-characterized females (19 African-American and 7 Caucasian) and 55 males (41 African-American and 14 Caucasian) with alcoholism agreed to participate for genetic testing. Genomic DNA was isolated from peripheral blood and CAG repeats determined by analyzing PCR amplified products using the polymorphic AR gene assay. The CAG repeat length for males and average CAG repeat length (CAGave) of both X chromosomes for females was correlated with raw scores from the Barratt Impulsivity Scale, version 11 and the Alcoholism Severity Scale. CAGave repeat lengths for Caucasian females [mean (SD)=18.1 (1.5)] were significantly longer than seen in African-American females [mean (SD)=16.5 (1.65); t = 5.3, p<0.03], but CAG repeat lengths in men did not differ by race [mean (SD)=15.4 (2.8)]. The average number of CAG repeats were significantly positively correlated (p<0.05) with impulsivity scores in females but not in males. The strongest relationship was observed for Caucasian females who showed a significant correlation for 4 first order factors (Motor Impulsiveness, Self-Control, Cognitive Complexity and Perseverance) and 2 second order factors (Motor Impulsiveness and Non planning Impulsiveness). Women with an average CAG repeat length of >18 representing the upper quartile of the repeat range showed significantly greater mean raw impulsivity scores. The CAG repeat length appeared to have less effect in African-American compared with Caucasian women possibly due to a shorter average repeat length. The results suggest that increasing CAG repeat length in females may contribute to psychopathology in alcoholism. Replicative studies with more females with alcoholism of both races are warranted.

Analysis of KDM5C defects associated with ARX Epilepsy-related mutants and evaluation of rescue strategies. L. Poeta1, A. Padula4, F. Fusco1, C. Shoubbridge3, G. Manganelli2, S. Filosa3, P. Collombat3, G. Fricourt4, M. Passafaro3, K. Helin1, L. Altucci1-2, S. Gustincich2, J. Gecz2, M.V. Ursini1, M.G. Milano1. 1) Institute of Genetics and Biophysics ‘Adriano Buzzati Traverso’, CNR, Naples, Italy; 2) Department of Paediatrics, University of Adelaide, South Australia, Australia; 3) Inserm U1091 Diabetes Genetics Team, Nice, France; 4) Inserm U1078 Laboratoire de Genétique Moléculaire et de Génétique Épidémiologique, Brest, France; 5) Institute of Neuroscience, CNR, Milan, Italy; 6) Centre for Epigenetics, University of Copenhagen, Copenhagen, Denmark; 7) Second University of Naples, Naples, Italy; 8) SISSA, Area Neuroscience, Trieste, Italy.

Malignant Epilepsy linked to Aristless-related homeobox (ARX) mutations presents severe pharmaco-resistant paediatric seizure. ARX is a crucial transcription factor regulating proliferation and migration of GABA-interneurons. We have recently identified a crucial epileptogenesis path, linking functionally ARX to another XLID/Epilepsy gene, Lysine-specific demethylase 5C (KDM5C). It encodes an H3K4me2/3 demethylase, which functioning is mediated by interactions with REST/NSRF, a master epigenetic hub critical for neuronal differentiation. We found that ARX Epilepsy-related mutations, which fall in PolyAlanine tracts or in the HD domain, cause a spectrum of functional damages of the ARX-KDM5C interaction, which severity depends on the type of alteration. In Arx KO embryonic brain and ES-oriented GABAergic neurons, a defective ARX-KDM5C-H3K4me3 path has been found in association with a mis-regulation of KDM5C Epilepsy-disease targets. Arx shRNA injections have been performed in mature neurons and endogenous ARX-KDM5C levels, spine density and morphology have been analyzed. We tested in vitro correction of KDM5C-H3K4me3 defects by exploiting three approaches: transcription factor targeting, gene knock-up and epigenetic modifications. To upregulate the endogenous KDM5C/KDM5C content, in a locus-specific manner, we tested KDM5C transcriptional induction by PHF8/ZNF711 stimulation and KDM5C translational increase by SINEUP method, aiming to balance the spectrum of KDM5C defects associated with the hypomorphic ARX PolyAlanine mutants. To achieve KDM5C-H3K4me3 path correction, we also screened a number of compounds targeting chromatin enzymes. We used an ES cell disease model, neuronally-differentiated Arx KO/Kdm5c-depleted ES cells, which show GABAergic abnormalities in association with a global increase of H3K4me3 signal. A strong compensation of Kdm5C/KDM5C downregulation has been obtained at crucial time-point of neuronal commitment. Although many other ARX targets could have important roles in the XLID/epilepsy phenotype, we believe that restoring or upregulating the expression of KDM5C gene or protein, ideally through endogenous physiological mechanisms, should accelerate the identification of alternative therapy to cure ARX Epilepsy phenotypes and many other neuropathologies with malignant seizure.
1400W Validation of a robust PCR-only assay for quantifying FMR1 CGG repeats on clinical samples. J.K. Moore1, M.J. Bashome2, G. Filipov3, K. Adler1, M.J. Fietz1, M. Schermer1. 1) Molecular Diagnostics, PerkinElmer, Inc, Waltham, MA; 2) Molecular Diagnostic Laboratory, Greenwood Genetic Center, Greenwood, SC.

Fragile X syndrome (FXS) is the most common form of inherited intellectual disability and is caused by an expansion of a CGG repeat in the 5’ end of the FMR1 gene. An accurate measurement of the repeat size is part of the classification and diagnosis of FXS and other Fragile X-associated disorders. To date, sizing of FMR1 trinucleotide repeats in the clinical laboratory has required hybridization methods or PCR-based methods with the manual preparation and labor intensive measurement of the number of CGG repeats by Southern blot. Our goal is to validate and establish a simple and robust PCR-only assay for quantification of CGG repeat alleles. A blinded set of 200 archived clinical DNA samples was analyzed by PCR amplification using a FragilEase™ kit (PerkinElmer). The repeat number was calculated from the fragment size measured on a low-cost capillary electrophoresis 2100 Bioanalyzer instrument (Agilent Technologies). The repeat numbers of those samples were interpreted by the use of standards with known repeats in each run. The sample set included both male and female individuals with normal (n=167), intermediate (n=10), premutation (n=11), and full mutation alleles (n=12). Calculated repeat sizes were compared to genotypes originally obtained by the current gold standard of PCR with fragment sizing on a capillary sequencer followed by Southern analysis of no-result PCR samples or single allele females. Results: All DNA samples were amplified successfully. The twelve full-mutation samples were successfully amplified with the largest allele size measured at over 653 repeats. We found the PCR amplifications in the premutation range to be more difficult to handle than those in the normal range. The turnaround time is within 6 hours and we can handle 10 samples in a single run. Conclusion: This new PCR-only method is capable of classifying all different FXS allele types correctly. Our PCR-only method allows Fragile X testing to be performed in a broader spectrum of clinical laboratories and is a cost-effective approach for diagnosis and potentially for population screening of the Fragile X syndrome.

1402F Applying NGS to familial ALS cohorts to identify novel genes. K.L. Williams1, J.A. Fifita1, G.A. Nicholson1,2, I.P. Blair1. 1) Australian School of Advanced Medicine, Macquarie University, Macquarie University, NSW, Australia; 2) Molecular Medicine Laboratory, Concord Hospital, New South Wales, Australia.

Amyotrophic lateral sclerosis (ALS) is a rapidly progressive neurodegenerative disorder. Familial ALS accounts for ~10% of ALS cases with the remainder being sporadic. ALS is genetically heterogeneous, with mutations in known genes explaining 60% of familial cases in Australia. We aim to combine genetic linkage analysis, exome sequencing, high-throughput genotyping and bioinformatic strategies to identify novel causative ALS genes in Australian ALS families. The 66 ALS families used in this study are categorised into Types 1, 2 and 3 based on sample availability. All families are negative for mutations in SOD1, TARDBP, FUS and the hexanucleotide repeat expansion in C9ORF72. Exome capture and sequencing (llumina TruSeq and HiSeq2000) was performed on 79 individuals from these families. Resultant linkage data was coupled with filtered exome data to identify 1, 2 and 4 variants respectively from the Type 1 families. These variants are currently being validated in extended patient and control cohorts, patient tissues, and by in vitro and in vivo functional studies. Type 2 families (n=3, sample n=10) have DNA samples from either two affected individuals plus a ‘married-in’ control or >2 affected individuals. Following exome sequencing and filtering, there were 18, 28 and 37 remaining potential causative variants across the Type 2 families. To reduce the number of variants, we are performing high-throughput iPLEX genotyping (Sequenom MassARRAY) in 700 matched population controls. Each potential candidate gene was also screened through exome data from family Type 3 index cases (n=60) to search for additional variants in these genes. SNPs are being developed to identify related individuals among family Type 3 index cases. If related individuals are identified they will be combined to create a Type 2 family and reduce the number of potential causative variants. We are implementing next generation sequencing in Type 2 families to test for previously unidentified ALS family types. Identifying novel ALS genes will give insights into the biological basis of motor neuron degeneration, allow development of new disease models and provide new targets for therapeutic development.

1401T AGG interruptions affect the risk of having a child with fragile X syndrome: a follow-up study. C. Yrigollen1, 2) D. Durbin-Johnson1, R. Hageman1, 3) L. Zhou2, E. Berry-Kravis1, P. Tassone3. 1) Biochemistry and Molecular Medicine, University of California Davis, Sacramento, CA; Public Health Sciences, University of California Davis, Davis, CA; 3) MIND Institute, University of California Davis, Sacramento, CA; 4) Pediatrics, University of California Davis, Sacramento, CA; 5) Neurological Science, Rush University, Chicago, IL.

AGG interruptions in the Fragile X Mental Retardation 1 (FMR1) CGG trinucleotide repeat are known to stabilize the estradiol 17β to estrogen receptor α. We have previously reported the presence of AGG interruptions within an FMR1 premutation allele reduces the risk of the allele expanding to a full mutation during maternal transmission particularly in the 60-90 CGG repeat range. It has also been shown that the instability of alleles of 45-69 trinucleotide repeats in the clinical laboratory has"
1404T
TDP43 interact with Pur α and rescue the neurodegeneration caused by Expanded Hexanucleotide GGGGCC Repeat. J. Li1,2, M. Poidevin1, W. Thomas1,2,4, P. Jin1. 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 2) Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, People’s Republic of China; 3) Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA; 4) Department of Veterans Affairs Medical Center, Atlanta, GA, USA.

Recently, an expansion of GGGGCC repeats in the first intron of C9orf72 was found to be a common cause of both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Previously, we found that Pur α interact with GGGGCC hexanucleotide repeat in vitro and in vivo in a sequence-specific fashion that is conserved between mammals and Drosophila. Here we have extended this work to investigate the role of TAR DNA-binding protein (TDP-43), the most common pathological inclusion in ALS/FTD, in riboGGG GCC(CCC)G repeat toxicity. To do this, we used NeuRo2a and Drosophila models of the expanded hexanucleotide repeat to show that the overexpression of the TDP43 is sufficient to rescue neurodegeneration. Next, we delineated the interaction between Pur α and TDP43. We show that Pur α interacts with TDP43 in an RNA independent manner that is dependent on the N-terminus of TDP43. Finally, we performed a Drosophila RNAi screen using TDP43 interacting protein to identify additional genetic interactors of Pur α / TDP43 complex. Our results suggest that the expanded GGGGCC repeat may interact with TDP43 indirectly through Pur α, causing the sequestration of TDP43 and other genetic interacting proteins from their normal function. This mechanism could potentially play a role in the pathogenesis of ALS/FTD illness.

1405F
A predictive model for the MAOA promoter VNTR and its application in public GWAS datasets. T. Wang1, A. Lu1, V. Rao1, S. de Jong1, R.M. Cantor1,2, R.A. Ophoff1,2,3,4,5. 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, California; 2) Department of Human Genetics, University of California Los Angeles, Los Angeles, California; 3) Department of Psychiatry, UMC Utrecht, The Netherlands.

Monoamine oxidase (MAO) A and B are adjacent genes on chromosome X that encode the main enzymes for neurotransmitter turnover. Many drugs used as treatment for clinical depression and anxiety inhibit the MAO enzymes, this suggests that they have an important role in psychiatric traits. Both MAOA and MAOB share a common promoter where a 30bp variable number tandem repeat (VNTR) was identified 1.2kb upstream the transcription start site of the MAOA. These alleles exist in the 2, 3, 3.5, 4, and 5 repeats (R), where 3R and 4R are the most common within the population. 2R, 3R, and 5R were shown to possess low activation of MAOA, whereas 3.5R and 4R were shown to have much higher levels of activation in vitro studies. However, the reported in vitro promoter high/low activity has poor correlation with in vivo studies examining MAOA or MAOB expression in post-mortem brain. Furthermore, these in vivo studies are underpowered to detect significant associations. We developed a predictive tool for the MAOA promoter VNTR based on an advanced machine learning method, multiscatter vertex discriminant analysis (VDA). We PCR genotyped the VNTR in 400 individuals of European descent, including 30 HapMap trios where there was SNP genotype information within 2MB of the MAOA promoter region. 300 individuals were used as a training dataset to build a VDA model using only overlapping SNPs in all datasets and searched by 110-fold cross validation to optimize the accuracy rate for predicting MAOA promoter VNTR. A test dataset of 87 individuals was used to validate the VDA model. This algorithm highlighted 5 SNPs, along with gender, that correctly classified the VNTR in 97% of our training dataset, and 96.7% in our validation dataset. With the predictor tool we are now able to study the MAOA promoter VNTR in available GWAS data sets of neuropsychiatric traits such as major depressive disorder, bipolar disorder and schizophrenia. The predictor tool enables us to establish whether some alleles are indeed high and low expressing (as reported in previous in vitro studies) in large expression data sets of human brain. Lastly, analysis is underway to test whether specific MAOA VNTR alleles affect monoamine neurotransmitter turnover by examining cerebrospinal fluid (CSF) monoamine metabolite levels in a large cohort of healthy controls.

1406W
MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington’s Disease Mice. S. Tome1,2, K. Manley1, J. Simard1,2, G.W. Clark1,2,3, M.S. Slean1,2, M. Swami4, P.F. Shelbourne4, E.R.M. Tiller4,5, D.G. Monkton4, A. Messer5, C.E. Pearson1,2,3, 1) inserm, Paris, France; 2) Genetics and Genome Biology, The Hospital for Sick Children, TMDT Building 101 College St., 15th Floor, Room 15-312 East Tower, Toronto, ON, M5G 1L7; 3) Wadsworth Center, New York State Dept. of Health, & Department of Biomedical Sciences, University at Albany, SUNY, Albany, NY 12208, USA; 4) Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow University Avenue, Glasgow G12 8QQ, UK; 5) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 6) Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada; 7) Campbell Family Institute for Cancer Research, Ontario Cancer Institute, University Health Network, Toronto, ON, Canada.

Expansions of trinucleotide CAG/CTG repeats in somatic tissues are thought to contribute to ongoing disease progression through a affected individual with Huntington’s disease or myotonic dystrophy. Broad ranges of repeat instability arise between individuals with expanded repeats, suggesting the existence of modifiers of repeat instability. Mice with expanded CAG/CTG repeats show variable levels of instability depending upon mouse strain. However, to date the genetic modifiers underlying these differences have not been identified. We show that in liver and striatum, the R6/1 Huntington’s disease (HD) (CAG)–100 transgene, when present in a congenic C57BL/6J (B6) background, incurred expansion-biased repeat mutations, whereas the repeat was stable in a congenic BALB/cByJ (CBy) background. Reciprocal congenic mice revealed the Msh3 gene as the determinant for the differences in repeat instability. Expansion-bias was observed in congenic mice homozygous for the B6 Msh3 gene on a CBy background, while the CAG tract was stabilized in congenics homozygous for the CBy Msh3 gene on a B6 background. The CAG stabilization was as dramatic as the genetic deficiency of Msh2. The B6 and CBy Msh3 genes had identical promoters but differed in coding regions, and showed strikingly different protein levels. B6 MSH3 variant protein is highly expressed and associated with CAG expansions, while the CBy MSH3 variant protein is expressed at background levels, and stabilizes the Dnmt1 protein, which is divergently transcribed from a promoter shared by the Msh3 gene, did not show varied levels between mouse strains. Thus, naturally occurring MSH3 protein polymorphisms are modifiers of CAG repeat instability, likely through variable MSH3 protein stability. Since evidence supports that somatic CAG instability is a modifier and predictor of disease, our data are consistent with the hypothesis that variable levels of CAG instability associated with polymorphisms of DNA repair genes may have prognostic implications for various repeat-associated diseases.
Posters: Psychiatric Genetics, Neurogenetics and Neurodegeneration

1407T

Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice:
genome-wide and candidate approaches. R. Mouro Pinto 1, E. Dragileva 1,
A. Kirby 2,3, A. Lloret 1, E. Lopez 1, J. St. Claire 1, G.B. Panigrahi 4,5, C. Hou 6,
K. Holloway 7, T. Gillis 1, J.R. Guide 1, P.E. Cohen 7, G-M. Li 6, C.E. Pearson 4,5,
M.J. Daly 2,3, V.C. Wheeler 1. 1) Center for Human Genetic Research, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts,
United States of America; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, United States of
America; 3) Program in Medical and Population Genetics, Broad Institute
of MIT and Harvard, Cambridge, Massachusetts, United States of America;
4) Program of Genetics and Genome Biology, The Hospital for Sick Children,
Toronto, Canada; 5) Department of Molecular Genetics, University of
Toronto, Toronto, Canada; 6) Graduate Center for Toxicology and Markey
Cancer Center, University of Kentucky College of Medicine, Lexington, Kentucky, United States of America; 7) Department of Biomedical Sciences,
Cornell University, Ithaca, New York, United States of America.
The Huntington’s disease gene (HTT) CAG repeat mutation undergoes
somatic expansion that correlates with pathogenesis. Modifiers of somatic
expansion may therefore provide routes for therapies targeting the underlying mutation. Huntington’s disease Hdh Q111 mice exhibit higher levels of
somatic HTT CAG expansion on a C57BL/6 genetic background
(B6.Hdh Q111) than on a 129 background (129.Hdh Q111). Linkage mapping
in (B6x129).Hdh Q111 F2 intercross animals identified a single quantitative
trait locus underlying the strain-specific difference in expansion in the striatum, implicating mismatch repair (MMR) gene Mlh1 as the most likely candidate modifier. Crossing B6.Hdh Q111 mice onto an Mlh1 null background
demonstrated that Mlh1 is essential for somatic CAG expansions and that
it is an enhancer of the HTT CAG pathogenic process in striatal neurons.
Hdh Q111 somatic expansion was also abolished in mice deficient in the Mlh3
gene, implicating MutLγ (MLH1/MLH3) complex as a key driver of somatic
expansion. Strikingly, the Mlh1 and Mlh3 genes encoding MMR effector
proteins were as critical to somatic expansion as Msh2 and Msh3 genes
that encode DNA mismatch binding proteins. The Mlh1 locus is highly polymorphic between B6 and 129 strains. While we were unable to detect any
difference in base-base mismatch or short slipped-repeat repair activity
between B6 and 129 MLH1 variants, repair efficiency was MLH1 dosedependent. MLH1 mRNA and protein levels were significantly decreased
in 129 mice compared to B6 mice, consistent with a dose-sensitive MLH1dependent DNA repair mechanism underlying the somatic expansion difference between these strains. Together, these data identify Mlh1 and Mlh3
as novel critical genetic modifiers of HTT CAG instability, point to Mlh1
genetic variation as the likely source of the instability difference in B6 and
129 strains and suggest that MLH1 protein levels play an important role in
driving of the efficiency of somatic expansions.

1408F

Small Molecule Ligand Distorts RNA G-Quadruplex Structure of the
Disease-Associated r(GGGGCC)n Repeat of the C9orf72 Gene and
Blocks Interaction of RNA-Binding Proteins. C.E. Pearson 1, 2, B. Zamiri 3,
K. Reddy 1, 3, R. Macgregor 3. 1) Genetics & Genome Biology, Hospital for
Sick Children, Toronto, Ontario, Canada; 2) Program of Molecular Genetics,
University of Toronto, Toronto, Ontario, Canada M5S 1A1; 3) Graduate
Department of Pharmaceutical Sciences, Leslie Dan Faculty of Pharmacy,
University of Toronto, Toronto, Ontario, Canada M5S 3M2.
Certain DNA and RNA sequences can form G-quadruplexes, which can
affect genetic instability, promoter activity, RNA splicing, RNA stability, and
neurite mRNA localization. Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) can be caused by expansion of a (GGGGCC)n
repeat in the C9orf72 gene. Mutant r(GGGGCC)n-containing transcripts
aggregate in nuclear foci possibly sequestering repeat-binding proteins like
hnRNPA1 - suggesting a toxic-RNA pathogenesis, as occurs in myotonic
dystrophy (DM1) and fragile X associated tremor ataxia (FXTAS). Furthermore, the C9orf72 repeat RNA and the FXTAS repeat RNA were recently
demonstrated to undergo the non-canonical repeat associated non-AUG
translation (RAN-translation) into pathologic peptide repeats in patient
brains, a process that is thought to depend upon RNA structure. RANtranslation was initially observed in tissues from DM1 and spinocerebellar
ataxia type 8 patients. We previously demonstrated that the C9orf72
r(GGGGCC)n RNA forms repeat tract length-dependent G-quadruplex structures that can be bound by the splicing factor ASF/SF2. Here we show
that a small molecule ligand, which can bind some G-quadruplex forming
sequences, can bind and distort the G-quadruplex of the r(GGGGCC)8
RNA, and this ablates interaction of either hnRNPA1 or ASF/SF2. These
findings provide proof-of-concept that nucleic acid binding small molecules
can distort the secondary structure of the C9orf72 repeat, which may beneficially disrupt protein interactions, which may ablate either protein-sequestration and/or RAN-translation into potentially toxic dipeptides. Disruption of
secondary structure of the C9orf72 RNA repeats may be a viable therapeutic
avenue for various diseases, as well as a means to test the role of RNA
structure upon RAN-translation.

307

1409W

Haplotype analyses in 15 Brazilian and Peruvian families with Spinocerebellar Ataxia type 10 (SCA10). M.L. Saraiva-Pereira 1,2,3,13,14 , T.C.
Gheno 1,2,3, G.V. Furtado 1,2,3,14, J.A.M. Saute 3, K.C. Donis 3, A.M.V. Fontanari 1,3, V.E. Emmel 1,2,3, J.L. Pedroso 4, O. Barsottini 4, C. Godeiro 5, H.
van der Linden 6, E.T. Pereira 7, W. Marques-Junior 8, R.M. Castilhos 3,14, I.
Alonso 9, J. Sequeiros 9, M. Cornejo-Olivas 10, P. Mazzetti 10, V.L. Torman 11,
L.B. Jardim 1,2,3,12,14, Rede Neurogenetica. 1) Lab. de Identificacao Genetica-Centro de Pesquisas, Hospital de Clinicas de Porto Alegre, Porto Alegre,
RS, Brazil; 2) Programa de Pós-Graduação em Genética e Biologia Molecular, UFRGS, Porto Alegre, Brazil; 3) Serviço de Genética Médica, HCPA,
Porto Alegre, Brazil; 4) Universidade Federal de São Paulo, São Paulo,
Brazil; 5) Universidade Federal do Rio Grande do Norte, Natal, Brazil; 6)
Centro de Reabilitação Dr. Henrique Santillo, Goiânia, Brazil; 7) Universidade Federal de Santa Catarina, Florianópolis, Brazil; 8) Universidade de
São Paulo, Ribeirão Preto, Brazil; 9) UnIGENe, Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal; 10) Centro de Investigación en Neurogenética, Instituto Nacional de Ciencias Neurologicas, Lima,
Peru; 11) Departamento de Estatística, UFRGS, Porto Alegre, Brazil; 12)
Departamento de Medicina Interna, UFRGS, Porto Alegre, Brazil; 13)
Departamento de Bioquímica, UFRGS, Porto Alegre, Brazil; 14) Instituto
de Genética Médica Populacional, INAGEMP.
Spinocerebellar ataxia type 10 (SCA10) is a neurodegenerative disease
characterized by cerebellar ataxia and seizures, due to expansion of a
pentanucleotide repeat (ATTCT) in the ATXN10 gene. To date, SCA10 has
only been reported in patients from Latin-American countries. The numerous
families of Amerindian descent with SCA10 mutations and the absence of
this variant in European countries suggest a founder effect for this mutation.
The objective of this study was to identify a previously described haplotype,
shared by 3 SCA10 Brazilian families, in additional patients from Brazil
and new cases from Peru. Repeat primed-PCR was performed to detect
expanded SCA10 alleles in Brazilian and Peruvian individuals with a SCA
of unknown cause. Haplotypes were constructed, based on polymorphic
markers within and outside the gene. We have identified 23 confirmed
SCA10 cases: 3 families have been described previously (Alonso et al,
2006; Almeida et al, 2009), while 12 families were newly identified, within
a large cohort of patients with ataxia. From those, 3 families were from
Peru, while the remaining 9 were from Brazil. The same polymorphic markers
and haplotypes were typed in 100 individuals (200 alleles) from the general
local population. A different haplotype distribution was found in SCA10
patients and the control group. A common haplotype, 19CGGC14, was
found in 11/13 of Brazilian and in 1/3 of Peruvian families. Haplotype
19CGGC16 was found in one Peruvian and one Brazilian families; the
remaining two families showed other haplotypes. Individuals with SCA10
mutations have been previously described in Brazil, whereas this is the first
report of SCA10 in Peru. We confirmed the presence of a consistently
recurrent intragenic haplotype, suggesting a common ancestry for most, if
not all patients. Supported by: CNPq, FIPE-HCPA, FAPERGS, INAGEMP, RIBERMOV.

1410T

Detailed audiological evaluation of a patient with xeroderma pigmentosum with neural degeneration. D. Mercer 1 , F. Tsien 2, A. Hurley 1. 1)
Department of Communication Disorders, Louisiana State University Health
Sciences Center, New Orleans, LA; 2) Department of Genetics, Stanley S.
Scott Cancer Center, Louisiana State University Health Sciences Center,
New Orleans, LA.
Xeroderma pigmentosum (XP) is a rare autosomal recessive condition
characterized by defects in DNA excision repair, leading to extreme sensitivity to ultraviolet light damage and increased susceptibility to melanomas
and carcinomas. Approximately 25% of XP patients also exhibit progressive
neural degeneration which includes mental deterioration, cortical thinning,
and sensorineural hearing loss. We describe a patient with XP complementation subtype D. The patient initially presented in our clinic at 7 years of age
after failing a hearing screening. At that time, he was found to have a
unilateral mild high-frequency sensorineural hearing loss with a configuration
suggestive of noise exposure. He also demonstrated central auditory processing deficits. He was re-evaluated at age 11 years, and his hearing loss
had progressed to a bilateral moderate sensorineural hearing loss. At age
13 years, the patient acquired type 1 diabetes. Genetic confirmation of XP
was obtained at 18 years of age. XP testing revealed compound heterozygous missense mutations in complementation group D, one of the XP subtypes associated with neural degeneration and sensorineural hearing loss.
Serial audiograms and electrophysiological testing have been obtained on
this patient over a course of 11 years. His hearing loss has gradually progressed to a bilateral precipitous sloping sensorineural hearing loss that is
mild in the low frequencies and severe-to-profound in the high frequencies.
Auditory evoked potentials showed deteriorating waveform morphology of
the auditory brainstem response and late evoked potentials (P300). The
change in P300 recordings may be due to cortical thinning, which has been
reported in XP with neural degeneration. In addition to neural degeneration
and peripheral neuropathy, the patient has been treated for claw-toe deformity. These findings will further characterize the phenotype of XP type D.


1411F Characterizing the aging human brain transcriptome. N. Pochel1,2,3, D. Borges-Rivera1, B. Haas1, S. Rajagopal1,2,3, J. Xu1,2,3, C. McCabe1,2,3, O. Gevaert1, G. Srivastava1,2,3, A. Regev1, J. Schneider1, D. Bennett1, P. De Jager1,2,3. 1) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women’s Hospital; 2) Harvard Medical School; 3) Broad Institute of MIT and Harvard; 4) Cancer Center for Systems Biology, Department of Radiology, Stanford University; 5) Rush Alzheimer’s Disease Center, Rush University Medical Center.

Aging in humans is often accompanied by cognitive decline which, in many cases, leads to dementia and a syndromic diagnosis of Alzheimer’s disease (AD). Our goal is to characterize the aging human brain transcriptome in order to gain mechanistic and functional insights into the neurobiological processes implicated in aging-related cognitive decline and Alzheimer’s disease.

We sequenced RNA from the dorsolateral prefrontal cortex (DLPFC) of 550 individuals from two prospective cohort studies of aging that include brain donation at the time of death: the religious order study and the memory and aging project. The subjects are all non-demented at the start of the study and, at the time of death, differ in the extent of aging-related cognitive decline, ranging from cognitively non-impaired individuals to mildly impaired individuals, and to individuals clinically diagnosed with Alzheimer’s disease at the time of death. In addition to the transcriptome data, we have genotype data on each individual and DNA methylation data from the same region of the brain in each subject, which inform our RNA analyses.

We applied ab initio and de novo transcriptome assembly approaches to systematically study the aging human brain transcriptome for differentially expressed genes and isoforms, novel alternative isoforms, as well as mutations and RNA editing events in expressed genes. We selected differentially expressed genes and isoforms, relying on the known annotated genes and isoforms, and we tested these for enrichment in known categories. We identified novel transcripts and (trans-)splicing variants, and to compare clinical characteristics between individuals stratified by the combination of aberration and expression/splicing changes.

1412W Mucolipidosis type IV: progressive gliosis, synaptic dysfunction and cognitive deficits in Mcoln1 knock-out mouse. Y. Gnislchuk1, S. Srin2, W. Ma3, N. Rudinsky3, M. Cotolle1, K. Stember1, E. Sapp3, A. Muzikansky1, M. Diffliga1, R. Betensky1, B. Hyman3, K. Sullivan1, E. Sapp3, S. Slaugenhaupt1, 1) Center for Human Genetics Research, Massachusetts General Hospital/Harvard Medical School, Boston, MA; 2) The Institute of Psychiatry, Kings College London, London, UK; 3) Mass General Institute for Neurodegenerative Disease, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. Mucolipidosis type IV (MLIV) is a lysosomal storage disease caused by severe cognitive impairment, motor decline and progressive loss of vision. It is caused by mutations in MCOLN1 gene, which encodes a lysosomal late endosomal non-selective cation channel TRPML1. The mechanism of how loss of TRPML1 leads to severe CNS pathology remains currently unknown and there is no therapy. Therefore, we are performing neuropathological examinations of an MLIV mouse model (Mcoln1 knock-out mouse), starting from very early stages and following disease progression. Using different experimental approaches such as behavioral testing, in vivo imaging, electrophysiology, histochemistry and electron microscopy, we describe structural and functional abnormalities associated with Mcoln1 loss in the mouse brain. Our results revealed that the first signs of motor and cognitive decline are seen at 2 months of age. These early behavioral phenotypes were accompanied by activation of astrocytes and microglia, as well as severe dysmyelination. Estimates of neuron numbers in the regions of the brain with the most astrocytosis and microgliosis revealed no neuronal loss at 2 and 3 months of age, indicating that the glial reaction is a consequence of neuronal death. Detailed examination of CA1 activity in the 3 month-old Mcoln1-/- brains in vivo showed no difference between knockouts and littermate controls, implying normal calcium homeostasis at this age. Evaluation of the synaptic function in the Schaeffer collateral pathway revealed elevated paired-pulse facilitation (PPF) and long-term potentiation (LTP) in Mcoln1-/- hippocampus. Changes in LTP were accompanied with impaired learning and memory in the fear conditioning test. Overall, observed changes in synaptic transmission were not a result of changes in either the CA1 or CA3 fields of the hippocampus, as determined by stereologic counts. However, in the CA1 stratum radiatum we observed changes in synaptic morphology in excitatory synapses, reduced thickness of myelin sheaths and accumulation of abundant aggregates of inclusion bodies. Altogether, here we report for the first time synaptic dysfunction and cognitive deficits associated with TRPML1 loss which could explain the clinical manifestations seen in MLIV patients; and the early involvement of glial cells in MLIV disease pathophysiology, suggesting a new avenue towards the development of therapies for this devastating disease.

1413T Identifying novel candidate genes for neurological diseases. K. Summers. Roslin Inst, Univ Edinburgh, Roslin, Midlothian, United Kingdom.

Over 80% of human genes are expressed in brain. Many genes are expressed in brain. Malignant brain syndromes (where groups of contiguous genes are deleted) almost always involve some form of intellectual disability and suggest that at least one in ten genes is critical for brain development. Within the mammalian genome there are a large number of genes that are not fully annotated and whose function is unknown. To identify such novel genes which are specific to the nervous system, the tool Biolayout Express2D was employed to cluster sets of genes with shared expression patterns across mouse neuronal cell types. This approach is model-free, since clusters are generated independent of known pathways or functions, and thus has the potential to find new pathway members. Published gene expression microarray data using the mouse MOE-430 platform (Affymetrix) were drawn from GEO-Datasets (GEO databases). Cell types and cell types analysed included a range of brain regions, spinal cord, glial cells and non-neuronal cells. Biolayout Express2D produced 187 coexpression clusters of more than 10 nodes (microarray probes). There were 16 clusters of genes (1027 probes representing 833 genes in total) expressed highly across the nervous system. Of these, 195 probes had no annotation and 151 had minimal annotation. Thus there was little information for one third of genes showing expression in brain and other neurological tissues. These genes included putative synaptic proteins, transcription factors of unknown target and other genes with no homology to any known protein domain. Knowledge of the other genes in the same cluster provides a functional annotation and identifies novel candidates for diseases affecting this function. Cluster 3 contained Env2, the gene for neuron specific enolase, indicating that this is a neuron-specific cluster, and genes for a number of known synaptic genes of minimal homology. For example, this cluster contained Dbc1 (Deleted in Bladder Cancer 1; OMIM602865), thought to be a tumour suppressor, and now implicated in neuronal/synaptic function. Another gene in this cluster is Susd4 (SusD Domain Containing 4; no OMIM entry), which was discovered by genome sequencing and is deleted in cases of autism and Fynys syndrome. Further analysis using knock-out mouse models, recombinant protein analysis and RNA interference will validate the role of the novel genes as candidates for diseases of the brain and nervous system.
The genetic architecture of schizophrenia: How do CNVs and polygenic scores contribute to disease risk? 


1) Medical Epidemiology and Bioeatastics, Karolinska Institute, Stockholm, Sweden; 2) Psychiatric and Neurodevelopmental Unit, Massachusetts General Hospital, Boston, MA; 3) Department of Psychiatry, University of California San Diego, La Jolla, CA; 4) Mount Sinai School of Medicine, New York, NY; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 6) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA.

Both rare CNVs and common SNPs contribute to the genetic risk for schizophrenia. Several specific CNV regions and an increased burden of large deletion CNVs have demonstrated associations with schizophrenia. A significantly higher polygenic risk score in subjects with schizophrenia has also been established, confirming that many common SNPs confer risk for this disorder as well. The relationships between these rare and common genetic risk factors have not been thoroughly investigated, and we sought to address the following questions: 1) Do cases with CNVs have lower polygenic risk scores compared to cases without CNVs? 2) Do cases with CNVs have higher polygenic risk scores compared to controls with CNVs? 3) Do controls with CNVs have lower polygenic risk scores than controls without CNVs?

We investigated the polygenic risk score differences within and between case and control groups by CNV carrier status using the Swedish Schizophrenia Consortium (N=4646) as the discovery sample to score the International Schizophrenia Consortium (ISC) (N=4921) subjects. Analyses were extended to CNV and GWAS data from the Psychiatric Genomics Consortium CNV and Schizophrenia groups. CNV carriers were defined by the two classes of CNVs conferring the greatest disease risks: 1) having one of 12 specific CNVs previously associated with schizophrenia or 2) carrying any large CNV deletion greater than 500 kb.

Within schizophrenia cases, CNV carrier did not demonstrate significantly lower risk scores than non-carriers. Cases with either class of CNV membership had similar scores when compared to control subjects carrying CNVs. However, in the ISC data, CNV carriers had lower polygenic risk scores than non-carriers. Cases with either class of CNV membership had higher polygenic risk scores than non-carriers. Cases with either class of CNV membership had lower risk scores than non-carriers. Cases with either class of CNV membership had lower polygenic risk scores than non-carriers.

The objective of this study was clearly explained and written informed consent was obtained from all subjects. The study was approved by the Ethical Committee of the Faculty of Medicine, the University of Tokyo. Results: We found no differences in copy number variant alleles in the three pairs of monozygotic twins. Conclusion: We have undertaken the high resolution copy number variation analysis in monozygotic twins discordant for schizophrenia and delusional disorder and identified no copy number variation differences within each monozygotic twin. CNV may not play a role in discordance of the development of psychosis in this present three monozygotic twin pairs.
1417F

Introduction: Clinical and experimental evidence have demonstrated that seizures activate the inflammatory response increasing the release of interleukin-1 beta and that this activation is age-regulated in the developing brain. In this study, we aimed to investigate the interleukin-1 beta (il1b) transcript response after seizures in immature zebrafish brain. Material and Methods: Zebrafish were maintained according to standard procedures and all experiments were approved by the Animal Ethics Committee/UNICAMP. Seven and 15 days post-fertilization (dpf) zebrafish larvae were separated in Seizure (SG) and Control (CG) groups. Animals from CG group were individually exposed to Pentylenetetrazol (PTZ) 15mM for 20min. Animals from CG were exposed to same handling condition, but in normal bath water. A pool of 20 heads was used to compose a single larvae sample (n=5 each group). One hour after PTZ exposure animals were cryoanesthetized and their heads were immediately isolated, frozen in N2(L) and total RNA extracted from Trizol (Invitrogen). Reverse transcription quantitative-PCR amplifications were carried out in triplicates with efla as endogenous control (TaqMan®, Applied Biosystems). The relative quantification (RQ) was calculated by the equation RQ=2-ΔΔCT and the statistical analysis was performed by Mann-Whitney test. Results: The mRNA profile of the il1b gene showed an up regulation in SG when compared with the CG in both 7 dpf ([CG1h 1.44 ± 0.34; SG1h 2.46 ± 0.17 (p=0.02)] and 15 dpf ([CG1h 1.96 ± 0.27; SG1h 3.63 ± 0.68 (p=0.02)]. The il1b transcript levels were 1.5 times higher in PTZ treated group. Conclusion: This is the first study investigating age-related response of il1b gene after seizure in immature zebrafish brain. Our results showed that seizures induced in immature zebrafish brain promote an age-specific cytokine expression. This study contributes for a characterization of zebrafish as an experimental model of epilepsies. Support: FAPESP, CNPq.

1418W
A corrective gene silencing by RNA interference to control over-expressed SNCA. M. Takahashi1, M. Suzuki2, N. Fujikake2, M. Murata2, K. Wada2, Y. Nagai2, H. Hohjoh1. 1) Dept. of Molecular Pharmacology, National Inst. of Neurosci., NCNP, Kodaira, Tokyo, Japan; 2) Dept. of Degenerative Neurological Diseases, National Inst. of Neurosci., NCNP, Kodaira, Tokyo, Japan; 3) National center Hospital, NCNP, Kodaira, Tokyo, Japan.

An over-expression of the wild-type alpha synuclein (SNCA) gene has the potential for developing Parkinson’s disease (PD). For reduction of the over-expressed SNCA, the gene silencing against the wild-type SNCA by RNA interference (RNAi) may be a possible strategy for treatment for such an SNCA-over-expressed PD. However, a conventional RNAi that thoroughly silences SNCA may be unsuitable for such a treatment, and previous studies suggested that the wild-type SNCA plays an important role in neuronal functions. In this study, we attempted to establish a new RNAi technique to control the level of the over-expressed SNCA to its normal level. Various siRNAs were designed and subjected to an in vitro assay to examine their RNAi activities. From the screening, siRNAs that conferred approximately a half of inhibition of the expression of wild-type SNCA were selected as potential candidates for the particular RNAi. To further assess the effects of such an RNAi treatment in vivo, we focused on Drosophila model of PD, and then we examined the siRNAs in cultured Drosophila S2 cells and slightly modified the siRNA sequences such that they could confer an appropriate RNAi activity in Drosophila cells. As for establishment of transgenic RNAi flies, the selected siRNA was converted into a shRNA-expression construct (plasmid) that is controlled by an eGFP-GAL4 driver for pan-neuronal expression. Using the resultant shRNA expression plasmid, we are currently generating transgenic flies, which express SNCA and the shRNA in neurons. We would like to present and discuss the data of the fly models as well as the selected siRNAs for controlling the over-expressed SNCA in the meeting.

1419T
Partitioning the Heritability of Tourette Syndrome and Obsessive Compulsive Disorder Reveals Differences in Genetic Architecture. L. Davis on behalf of the Tourette Syndrome Association International Consortium for Genetics (TSAICG). Section Genetic Medicine, The University of Chicago, Chicago, IL.

Despite the completion of the first genome-wide association studies in Tourette Syndrome (TS) and obsessive compulsive disorder (OCD), very little is known about the genetic architecture of these two phenotypically related early-onset neuropsychiatric disorders. The direct estimation of heritability from genome-wide common variant data as implemented in the program Genome-wide Complex Trait Analysis (GCTA) has provided a means to quantify and partition heritability attributable to all interrogated variants or to specific variant subsets of interest. We conducted multiple partitioning analyses to identify genomic elements that concentrate TS and OCD heritability. We partitioned by chromosomal, MAF bins, and by annotation of variants that regulate gene expression in the brain. In addition, we assessed heritability for early onset and adult onset OCD and finally tested for genetic overlap between TS and OCD. After extensive quality control, we proceeded with analysis on a final data set of 617 TS cases and 4,116 TS controls genotyped on 393,387 SNPs, as well as 1,061 OCD cases and 4,236 OCD controls genotyped on 373,846 SNPs. Our analysis yielded a heritability point estimate of 0.58 (se=0.09, p=5.64e-12) for TS, and 0.37 (se=0.07, p=1.5e-07) for OCD. Among other notable results, we found that SNPs with a minor allele frequency of less than 5% accounted for 21% of the TS heritability and 0% of the OCD heritability. We discovered a disproportionately large contribution to OCD heritability originating from chromosome 15. Additionally, results showed that pairwise eQTLs accounted for significantly more TS heritability (p=5.36e-46) and OCD heritability (p=2.80e-16), and cerebellum eQTLs accounted for significantly more TS heritability (p=4.02e-15) and OCD heritability (p=1.37e-14) than expected based on the number of SNPs tested under a uniform distribution model. Finally, we found a bivariate genetic correlation of 0.41 (SE=0.15) between TS and OCD. These findings suggest that 1) very little, if any, heritability is truly missing (i.e., unassayed) from TS and OCD GWAS studies of common variation and 2) shared environment does not result in excessive bias in twin and family studies of TS and OCD. The results further indicate that while there is some overlap between these two phenotypically-related neuropsychiatric disorders, the two disorders have distinct genetic architectures. We present genomic characterization of both phenotypes.

1420F
Association of DLG4 haplotype with increased risk of schizophrenia. S. Balan1, K. Yamada2, E. Hattori3, Y. Iwayama1, T. Toyota1, T. Oishi3, M. Maekawa1, M. Toyoshima1, Y. Iwata3, K. Suzuki1, M. Kikuchi1, T. Yoshikawa1. 1) Laboratory for Molecular Psychiatry, RIKEN-Brain Science Institute, Saitama, Japan; 2) Department of Psychiatry and Neurology, Hama-matsu University School of Medicine, Shizuoka, Japan; 3) Department of Psychiatry and Neurobiology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan.

The post-synaptic density (PSD) of glutamatergic synapses are characterized by the expression of an array of proteins critical for maintaining synaptic dynamics. Alteration of protein expression levels in this matrix is a marked phenomenon of neuropsychiatric disorders including schizophrenia, where cognitive functions are impaired. To query the genetic predisposition of the genes expressed in the PSD with schizophrenia, a family-based association analysis of genetic variants in PSD genes viz; DLG4, DLG1, PICK1 and MOD2, was performed, in Japanese pedigree samples (124 pedigrees, n = 376 subjects). A significant association of the variant rs17203261 in the DLG4 gene was observed, with preferential transmission of the C allele (p = 0.02), although significance disappeared after correction for multiple testing. However, this is confounded by a potential association of the variant with the Chinese schizophrenia cohort (293 pedigrees, n = 1163 subjects) or in a Japanese case-control sample (n = 4182 subjects). Further the DLG4 expression levels between postmortem brain samples from schizophrenia patients showed no significant changes from controls. Interestingly, a five marker haplotype in DLG4, involving rs22442449, rs17203281, rs590200, rs222853 and rs222897, was enriched in a population specific manner, where the sequences A-C-C-C-A and G-C-C-C-A accumulated in Japanese (p = 0.0009) and Chinese (p = 0.0007) schizophrenia pedigrees samples, respectively. None of the variants in other examined candidate genes showed any significant association in these samples. The present study highlights a putative role for DLG4 in schizophrenia pathogenesis, evidenced by haplotype association, and warrants further dense screening for variants within these haplotypes.
Children with 22q11.2 Deletion Syndrome (22q11.2 DS) are at elevated risk for Autism Spectrum Disorder (ASD), with 20% of individuals receiving a diagnosis. The medical and psychiatric manifestations of 22q11.2 DS are quite variable between individuals, despite most having involvement of the same 3 Mb region encompassing ~45 genes. Genetic modifiers for the risk of ASD in 22q11.2 DS as well as all other forms of syndromic ASD are poorly understood. We previously showed that copy number variants in the metabotropic glutamate receptor (mGluR) network are associated with an elevated risk of syndromic ASD and are uncommon in typically developing children. We hypothesized that alterations in the mGluR network outside of the deleted region may provide a ‘second hit’, conferring greater risk for development of ASD. The purpose of this study was to determine whether alterations in the mGluR network outside of 22q11.2 are associated with an increased risk of ASD in children with 22q11.2 DS. METHODS: DNA from 75 children with 22q11.2 DS were selected, including children with clinical diagnosis of ASD (n=25) and without ASD (n=50). High density microarray and custom gene chip were used to determine whether children had copy number variants in genes in the mGluR network, including those in the context of another genetic syndrome. We previously demonstrated alterations in the mGluR network are enriched in children with many forms of syndromic ASD. All individuals with classic 22q11.2 DS are missing one copy of RANBP1, an mGluR network gene located within the deleted region. Decreased expression of RANBP1 is also supported by results from our previous study to thalidomide and valproic acid, which are both associated with increased risk of ASD. While all individuals have one missing copy of RANBP1, the rate of ASD in 22q11.2 DS is only 20%, similar to the rate in younger siblings of children with ASD. This suggests that additional factors contribute to risk of ASD in 22q11.2. Our data suggest that CNN alterations in the mGluR network may be one important factor in determining which children with 22q11.2 DS will develop ASD.
Neural celluar models of 22q11.2 DS exhibit disruptions to the miRNA regulatory pathway and may increase the rate of schizophrenia in individuals with 22q11.2 DS. W. Manley, M.R. Ababon, M.P. Moreau, P.G. Matteson, J.H. Millonig, L.M. Brzustowicz. 1) Department of Genetics, Rutgers University, Piscataway, New Jersey 08854-8095, USA; 2) Center for Advanced Biotechnology and Medicine, Piscataway, NJ 08854, USA; 3) UCRD, Rutgers University, Piscataway, New Jersey 08854-8095, USA.

Schizophrenia is a complex and poorly understood disease caused by the interplay of many environmental and genetic factors. The 22q11.2 deletion syndrome is a disorder caused by the microdeletion of part of chromosome 22 leading to a 25-fold greater chance of developing schizophrenia in affected individuals versus the general population. The missing 22q11.2 region contains the DGCR8 gene, which is required for the initial step of miRNA biogenesis. However, the 22q11.2 deletion itself is not sufficient to cause schizophrenia, since 75% of individuals with this deletion do not develop the disease. We hypothesize that the 22q11.2 deletion increases the risk of schizophrenia through alterations in miRNA regulatory networks via deletion of multiple miRNAs. These miRNAs may serve as a protective buffer against the accumulation of deleterious mutations at other schizophrenia risk loci. We have profiled miRNAs in multiple cell lines from individuals with 22q11.2 that are haploinsufficient for DGCR8. Additionally, we have also developed neuronal human cellular models of disruptions to the miRNA regulatory network caused as a result of DGCR8 deficiency. The neuroblastoma cell line SH-SY5Y was transfected via electroporation with GIPZ lentiviral shRNAmir constructs specific for DGCR8 (Open Biosystems). The GIPZ constructs were prepared using PureYield Plasmid Maxi Prep System (Promega) and transfection with lipofectamine RNAiMAX. We selected constructs contained genes for GFP and puromycin resistance. The levels of DGCR8 gene expression was quantified using the 7900HT Fast Real-Time PCR System (Applied Biosystems) along with Taqman Gene Expression Assay. We used WT mouse cells and DGCR8 reduction human cells to determine if the decrease in DGCR8 expression could be involved in the elevated risk of developing schizophrenia in individuals with 22q11.2 deletion syndrome versus the general population. We have also performed RNA-seq using Illumina TruSeq protocols to characterize differences at the RNA level between DGCR8 deficient cells and otherwise healthy control cells. Future studies will focus on determining the biological targets of these miRNAs using miRNA target prediction software.

A case of monozygotic female triplets diagnosed with autism: comprehensive genomic analyses. Z. Talebizadeh, S. Soden, A. Gadashova. Pediatrics, Children's Mercy Hospital, Kansas City, MO.

Twin study design is a classical approach commonly used in many genetic studies; however, it enhances the power of evidence for detecting genetic causes of complex disorders. Autism is a common early onset neurodevelopmental disorder belonging to a group of conditions known as autism spectrum disorders (ASDs). Although there is strong evidence for genetic involvement in susceptibility to ASD, the etiology of most cases remains unknown. We present a case of female triplets diagnosed with autism, and discuss results of their genomic data analyses, including genome-wide copy number variations (CNVs) and expression profiling. No pathogenic CNVs were detected in the triplets’ peripheral blood samples. Genome-wide microRNA and exon array profiling were performed, and compared with three unaffected females as controls, to examine the potential abnormalities at the noncoding or alternative splicing levels. The exon arrays were analyzed at both exon and gene levels. One hundred thirty five genes showed differential exon expression in the triplets, compared with controls. By correlating expression levels of microRNAs with their predicted target mRNAs, differential expression was detected in 51 genes (i.e., inversely correlated with microRNAs). Pathway analysis using the 135 potentially alternatively spliced genes and the 51 genes obtained from microRNAs and mRNAs correlation showed enrichment in biologically relevant pathways. All three female triplets had skewed patterns of X chromosome inactivation. This, gene ontology enrichment analysis found significant biological and molecular processes shared among meditators’ higher state of consciousness. The number of differently expressed genes as well as high proportion of genes themselves differed between meditators. Despite this, gene ontology enrichment analysis found significant biological and molecular processes shared among meditators’ higher state of consciousness.

Does Presymptomatic testing changes Age at ONset: a prospective study in Huntington disease (PAON study). A. Durr, L.M. Brzustowicz, J. Boni, J. Dreo, J. Feingold, M. Gargulo. 1) APHP Genetic Department, Groupe Hospitalier Pitié-Salpêtrière Charles- Foix, Paris France Paris, France; 2) AP-HP, Groupe Hospitalier Pitié-Salpêtrière Charles-Foix, Department of Biostatistics and Medical Informatics, Paris, France; 3) ICM (Brain and Spine Institute), Groupe Hospitalier Pitié-Salpêtrière Charles-Foix, Paris France; 4) Inserm, UMR_S975, CRICM, F-75013, Paris, France, Groupe Hospitalier Pitié-Salpêtrière Charles-Foix, Paris France; 5) UPMC Univ Paris 06, UMR_S975, F-75013, Paris, France.

In PAON, we studied prospectively the impact of presymptomatic testing (PT) on the onset of symptoms in Huntington disease (HD). The goal of our study was to estimate a possible test-effect on age at onset in carriers of the mutation who choose to know their genetic status. Does knowledge about being a carrier of the pathological mutation changes age at onset or onset modalities? In Paris, at the Salpêtrière University Hospital - Genetic Department, 1634 at-risk persons for HD requested genetic testing since 1992 and entered multistep and multidisciplinary care and counseling. We went back to those who received a test result before 2009, in order that their age at examination was reaching the mean expected onset of 40 years. Among them, 463 were non carriers and 302 carriers of an abnormal expanded CAG repeat. Detailed follow up examination was available for 208 individuals (69%). We interviewed and examined additional 62 persons (PAON cohort), using a self-administered questionnaires (to explore their carrier condition) and the UHDRS evaluation score. We estimated age at onset with the Langbehn formula (personal communication of Doug Langbehn formula underestimates onset. Importantly, for carriers of the carrier condition) and the UHDRS evaluation score. We estimated age at onset with the Langbehn formula (personal communication of Doug Langbehn, 2004 Apr;65(4):267–77). Among the PAON cohort, there were 35 affected with UHDRS >5 (mean 18.3 +/- 11.9) and 27 unaffected (mean 2.9 +/- 1.8). Surprisingly, age at onset observed was 6.0 +/- 8.7 years earlier than onset calculated. Among those followed more regularly, 60 were included (mean 28.1 +/- 18.0) and had age at onset 5.2 +/- 7.2 years earlier than calculated. This did not differ significantly from a retrospective cohort of patients who did not have PT (n=522) where age at onset was 3.6 +/- 9.6 years earlier (p= 0.13 and p= 0.12, respectively). In conclusion, a presymptomatic testing was associated with an early onset, potentially leading to better overall prognosis. Future studies are needed to confirm these results.
Differences in gene expression and DNA methylation in drug-naïve first-episode psychosis patients. V.K. Ota, C.S. Noto, A. Gadelha, M.L. Santoro, B.B. Ortiz, R.S. Silhano, E.S. Gouveia, P.N. Silva, C.G. Olmos, L.M.N. Spindola, E.H.S. Andrade, M.I. Melaragno, M.A.C. Smith, S.W. Han, Q. Cordeiro, R.A. Bressan, S.I. Belangero. 1) Morphology and Genetics, UNIFESP, Sao Paulo, Brazil; 2) Psychiatry, UNIFESP, Sao Paulo, Brazil; 3) Interdisciplinary Center for Gene Therapy, UNIFESP, Sao Paulo, Brazil; 4) Psychiatry, ISCMSP, Sao Paulo, Brazil.

Schizophrenia is a severe mental health disorder with a high heritability. The study of gene expression levels in blood of patients at the beginning of the disease, such as first-episode of psychosis (FEP) may be useful to detect changes in gene expression despite treatment effects. In this study we aimed to analyze gene expression in whole blood, comparing: a) drug-naive FEP patients and healthy subjects; and b) drug-naive FEP patients before and after treatment with antipsychotic drugs. Also, we investigated if those differentially expressed genes were regulated by DNA methylation.

All the patients (n = 38) were evaluated by a trained psychiatrist, twice (at admission and 8 weeks after beginning antipsychotic treatment), and the healthy controls (n - 38) were also evaluated to exclude any psychiatric disorder. Whole blood was collected from each participant during clinical assessments. Expression levels of 40 genes related to neurotransmission and neurodevelopment were quantified with a customized RT2 Profiler™ PCR Array, which is based on SYBR Green detection of cDNA amplification. For methylation analysis, bisulfite sequencing was performed. For data analysis, we compared 2-Dct or DNA methylation percentage values using t-test (FEP patients x healthy controls) or paired t-test (before x after treatment). Significant downregulation of GCH1 gene was observed (Fold regulation (FR) = -1.34, p=0.007) comparing FEP and controls. Downregulation of GABRR2 (FR = 1.33, p=0.01) and upregulation of GCHFR (FR = 1.18, p=0.008) were found in FEP after treatment. Also, a significant hypermethylation in GCH1 was detected in FEP comparing to healthy controls (p=0.034). GCH1 contains for GTP cyclohydrolase I, an enzyme involved in the synthesis of BH4, which is an essential cofactor for tyrosine, serotonin and L-Dopa. Moreover, its expression is regulated by GCHFR, which seemed to be upregulated after treatment with antipsychotic drugs. Also, GABRR2, which is a GABA receptor gene, seems to be dysregulated after treatment with antipsychotic drugs. Therefore, GCH1 may have a role in the genesis of psychosis whereas GABRR2 and GCHFR may be related to the treatment leading towards a better understanding of illness.

Funding for this study was provided by FAPESP 2010/19176-3; 2010/08968-6 and 2011/50740-5.


FOXP2 is the first key gene involved in the development of human language. FOXP2 protein is a highly conserved transcription factor in any animals. Only two amino acids are different between human and chimpanzee FOXP2 proteins, and this difference was supposed to be associated with neuron development for language skill between the two species. Previous study demonstrated that the FOXP2 gene was expressed not only in neurons but also in multiple tissues.

The aim of the present study is to identify genes targeted by FOXP2. We first generated various transgenes (Tg) in hek293-derived cells, using the flip-in recombinase system for the gene integration at the identical chromosome region and the Tet-ON/OFF systems to minimize transfection artifacts. The Tg included human FOXP2 (Hum), a human FOXP2 isoform (Iso), chimpanzee Foxp2 (Chimp), or a negative control (CAT). This iso was encoded by an alternative transcript with the complete open reading frame for a human FOXP2 with extra 25 amino acids, and was confirmed to be expressed in multiple human adult tissues, such as heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas, but not in any mouse tissues. We analyzed expression of the FOXP2 from the transgenes (FOXP2-Tg) in Tet-ON or in Tet-OFF cells. The expression of FOXP2-Tg in each Tet-ON cells was higher than in Tet-OFF cells, and FOXP2-Tg in Tet-ON cells were 5-10 times more highly expressed than that in CAT. Using these cell lines, we performed microarray screening to pick up altered expression pattern among these 8 groups. Then, we confirmed the expression pattern again in GOTO cells, which is initially used as a negative control, with stable and random integration of each Tg, using real time PCR. Consequently, results of the microarray analysis and real time PCR suggested several candidate target genes: CRYAB, GABA receptor genes, UBE3A, GFRA and EDRNBR. The CRYAB gene was upregulated in all Tet-ON cells, and the gene was more highly expressed in Hum and Chimp Tet-ON than the CAT Tet-ON cells. Similar expression patterns were confirmed with Tg in GOTO cells. Interestingly, genes for some GABA receptors and the UBE3A gene were also confirmed with the expression pattern of these Tg.

The present study shows that the FOXP2 gene and its isoform may be a good target gene for neuropsychiatric disorders as well as language development.
Reference-free quality assessment and complexity estimation of next-generation sequencing data. S. Y. Anvar1-3, L. Khachatryan1, M. Vermaat1-3, M. van Galen1, I. Pulyakhina1, Y. Anyurek1, K. Krajeveld2, P.A.C. 't Hoen1, J.T. den Dunnen1-2, P. de Knijff2, J.F.J. Laros1-3, 1) Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Leiden Genome Technology Center, Leiden University Medical Center, Leiden, Netherlands; 3) Netherlands Bioinformatics Center, Nijmegen, Netherlands; 4) University of Applied Sciences, Leiden, Netherlands.

Current methodologies for determining the quality and complexity of next-generation sequencing (NGS) data heavily rely on alignment-based quality measures. However, reference-gene sequences are not always available and genomic information about complex regions of the genome such as large duplications or structural rearrangements are likely to be overlooked by alignment-dependent approaches. To address these limitations, we have developed a new methodology (kMer) that is independent of a reference sequence in determining the quality of NGS data and allows for pairwise comparison across a series of samples. kMer provides various tools to systematically and dynamically assess the NGS data after generating k-mer profiles. In addition, kMer provides a framework to estimate the complexity of the NGS data library as well as estimating if the sample is sequenced in sufficient depth. We have applied kMer on four sets of NGS data that consist of 59 targeted resequencing, 43 whole-exomes, 49 full-genomes, and 665 RNA-Seq samples. In our test set, kMer could precisely detect and separate samples based on various technical variations that were introduced during sample prep or sequencing such as high duplication rate, low capture performance, differing capture protocols, and high amount of library chimaeras. Notably, some of the aforementioned artefacts could only be characterised after vigorous quality assessments and have otherwise been missed by standard NGS QC tools. These often hidden artefacts undermine the potential application of NGS in clinical diagnostics as they may result in obscured downstream analysis. In addition, we show that kMer allows for estimating the complexity of NGS data, a vital property of NGS in diverse studies ranging from de novo assembly applications to detection of a shift in abundance in a series of microbiomes. In particular, we show that the complexity of NGS data is reflected on the distribution of k-mer profiles, with the resolution of separating differing scenarios. In conclusion, kMer can be used to estimate the complexity of NGS data and to detect unexpected artefacts. The application of kMer to real-world data will be the subject of future studies.

1343F

Genomic structural variation (SV) represents a significant, yet poorly understood contribution to an individual’s genetic makeup. Advanced next-generation sequencing technologies are widely used to discover such variations, but there is no single SV detection algorithm that is considered a community standard. In response to this challenge, we developed a memory-efficient algorithm called SoftSearch, for discovering genomic rearrangements in Illumina paired-end next-generation sequencing data. SoftSearch simultaneously utilizes multiple strategies for detecting SV including split-read, discordant read-pair, and unmapped pairs. Split-reads and read pairs that are co-localized and supporting the same directionality are used to refine the breakpoints and classify them according to their type, e.g. large insertions-deletions, inversions, duplications and balanced or unbalanced inter-chromosomal translocations. SoftSearch outputs predicted structural variants in universally accepted VCF format. After extensive comparisons we attempted to recall known structural variants that have previously been validated in the HamMap NA12878 sample. SoftSearch recalled the most true positive results (n=660), 4 times more variants than BreakDancer (n=165). DELLY identified a comparable number of variants (n=654) variants, while CREST had the fewest (n=505). CREST and DELLY took the most amount of time to complete with more than 230 and 194 hours, respectively. The memory usage was also highest with these tools (5.2, and 7.1GB, respectively). BreakDancer was faster and used less memory requiring 48 hours and 1.5GB memory. Strikingly, SoftSearch completed the analysis in 20 hours using only 0.7GB memory. Only a small number of soft-masked bases from split reads and a few discordant read-pairs are necessary to identify an SV, which on their own would not be sufficient to make an SV call. Key features of SoftSearch are 1) not requiring secondary (or exhaustive primary) alignment, 2) portability into established sequencing workflows, and 3) is applicable to any DNA-sequencing experiment (e.g. whole genome, exome, custom capture, etc.). We show that SoftSearch can identify more true SVs with less evidence, including clinically relevant SVs in the BRCA2 gene missed by other SV tools while offering significantly improved overall performance.

1343W
Thousands Genomes And HLA Typing By NGS: Hidden Treasures In Public Short Read Data. A. Barrow1, E. Major1, K. Rigo1, S. Juhos1, T. Hague1, P. Pournara1. 1) Omixon Biocomputing, Budapest, Hungary; 2) University of California, San Francisco, US.

One of the important goals of the 1000 Genomes project was to find common mutation in diverse populations with the help of next generation sequencing. Earlier, the HapMap project with similar goals made it possible not only to map frequent mutations but there are already publications about sequencing based HLA typing using NGS. We are presenting a sufficiently fast algorithm using 1KG Illumina data to obtain HLA types for HLA-A, B, C, DRB1 and DQB1 genes. For validation the results of Sanger capillary sequencing based HLA typing was used for over thousand Coriell samples. According to our results it is possible to determine the correct HLA types from public 1KG whole-exome Illumina data with 90% or higher concordance if proper quality check measures are applied. We are also presenting the possible causes of mistypings. The method opens perspectives for typing other systems like MICA, MICB and KIR.
A consolidated genotype quality control reporting pipeline. J. Boston, Y. Bradford, J. Haines, W. Bush. Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Quality control is a critical part of any genetic analysis and must be performed in a reproducible, well-documented, and interactive way. To improve the throughput and reliability of this QC process, we developed a Clojure-based pipeline for generating quality control reports and plots in a systematic fashion. From a command-line interface, analysts submit data files to a centralized server and specify options for QC procedures. These include typical operations such as sex concordance, genotype call rate, sample efficiency, allele frequency distributions, Hardy-Weinberg tests, identity by descent calculations for relatedness, sample concordance checks, and tests for Mendelian inconsistencies. QC jobs are submitted to a scheduled computing cluster through portable batch system (PBS) scripts. Once completed, results from all operations are compiled into a single report of relevant information, complete with appropriate plots, eliminating the possibility of data transcription (through cut and paste operations) or other data processing errors. Because genotype QC is an iterative process, changes to the filtering process can be made after examining patterns in the original report, simply by reissuing the PBS script with new parameter settings. As such, the entire QC process is much more standardized and less prone to error. This consolidated workflow also improves and facilitates communication between analysts, study designers, and technicians about QC-related issues through a standardized report with multiple visualizations. While the pipeline was implemented for processing of genome-wide association data, the application is also applicable to next-generation sequence data.

Variome-based identification of Crohn’s disease predisposition. Y. Bromberg. Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ, USA.

Crohn’s disease (CD) is an auto-inflammatory disorder and one of the main subtypes of inflammatory bowel disease. CD affects over half a million people in North America alone. Tens of thousands of new CD cases are diagnosed every year, mostly in young adults in their twenties and early thirties. Human and animal models of CD illustrate strong familial patterns of disease predisposition, suggesting interplay of genetic and/or environmental causes. Recent GWAS studies highlight key CD pathogenic pathways, but the discovered susceptibility loci account for less than 25% of disease heritability. Rather than the result of a single very severe malfunction, most Crohn’s manifestations are thus likely due to ‘unlucky’ summations of many variants of individually small functional effects, as is the case with many other complex diseases. We have started building VariAD (Variome Analysis of Disease) - a novel computational tool for annotating an individual’s CD disease predisposition on the basis of a combination of genome variants (individual variome). Disease associated variants often act by disrupting the sequence and the molecular functionality of the affected gene-product. Our method assumes that a genome-specific set of functionally significant coding single nucleotide polymorphisms (cSNPs) in CD genes is sufficient to describe an individual’s CD susceptibility. We collected an expanded set of CD genes by mining scientific literature for phenotypic descriptors of CD. We annotate the ‘broken-ness’ of extracted disease genes based on their corresponding variants, taking into account the specific locus genotypes. The genome-specific combination of affected disease-genes drives the prototype CD prognosis engine. This research represents a unique way of looking at the variation in a single human genome, informed by molecular mechanisms of disease and achieves over 75% accuracy of prediction on the two data sets tested (122 case/control individuals). VariAD will motivate new testable hypothesis regarding the biological mechanisms of CD and provide a means for earlier prognosis, diagnosis and development of better treatments.

Clinical exome quality assurance through comprehensive coverage analysis improves the utility of exome sequencing. C.J. Buhay1, Q. Wang2, M. Wang1, Y. Han1, H. Dinh1, H. Dodapataneni1, Y. Yang2, Y. Ding2, M. Bainbridge3, E. Boerwinkle4, J. Reid1, D. Muzny5, R. Gibbs1. 1) Baylor College of Medicine - Human Genome Sequencing Center, Houston, TX; 2) Baylor College of Medicine - Whole Genome Sequencing Laboratory, Houston, TX.

It is important for whole exome sequencing in a clinical setting to continually increase probability of diagnosis and improve clinical utility. Comprehensive coverage improvements of clinical exomes occur in three phases: annotation of targets of interest in the exome, automatic detection of inadequately covered exons, and development of methods to rescue clinically significant low coverage regions. To better characterize coverage across one popular whole exome design (VCFom2.1), targeted gene regions across 34 clinical exome samples were scrutinized. We focused on the performance of relevant diagnostic genes across the HGMD, Genetest, and COSMIC lists. Results show that >90% of genes in the list have a base-depth coverage of 20X or better in our clinical samples. Leveraging the coverage tools developed in the process, we have implemented the Exome Coverage and Identification (EXCID) Report in our research pipeline. This workflow assesses exome coverage, annotates target regions with gene and exon coverage depth, as well as reports inadequately (<20X) covered exonic regions for every exome coming through the production pipeline. This tool allows investigators to efficiently assess performance across regions of interest. As part of a pilot project in the DNAnexus cloud platform, we are aggregating coverage plots from >5000 whole exome samples from our production pipeline. The implications of analyses at this scale are twofold: unsurpassed granularity of aggregate base coverage that leads to precise targeting and rescue of inadequately covered bases.
Pathogenicity prediction of genomic variants using the gene-level frequency of variation in asymptomatic individuals. C.A. Cassa, D.M. Jordan, M.S. Lebo, S.R. Sunyaev. 1) Division of Genetics, Harvard Medical School/Brigham and Women's Hospital, Boston, MA; 2) Program in Biophysics, Harvard University, Cambridge, MA; 3) Lab for Molecular Medicine, Partners Center for Personalized Genetic Medicine, Cambridge, MA.

The clinical importance of novel genomic variation is often characterized using in silico techniques, which rely on the exclusion of common variation or inclusion of variants with deleterious effects predicted using evolutionary and functional considerations [1, 2]. We extend these techniques by considering the expression patterns in each gene, genome-wide, to predict the robustness of each gene to variation. Genes vary greatly in the number of missense and nonsense variants present in asymptomatic individuals, which provides important context about selective pressure and significance of new variants that are observed. Using data from asymptomatic individuals in the Exome Sequencing Project (N=6503), we calculate the expected numbers of heterozygous, homozygous rare, and compound heterozygous variants in each gene. We then use these data to rank-order genes, for different variant types (missense, nonsense) and minor allele frequency bins, across two populations. We then use these expected numbers of variants per gene to annotate a known list of variants that have pathogenicity classifications from a clinical genetics laboratory (Lab for Molecular Medicine). We use these variants to train a general pathogenicity classifier, using the random forest method with 10 trees and 5 attributes per tree (out of 24 total attributes) [3]. We trained and tested this classifier using 10-fold cross validation on missense variants that were classified as either pathogenic, likely pathogenic, likely benign or benign (N=5737). The cross-validation results indicate that the classifier is able to distinguish between the benign or pathogenic classes. This demonstrates that the gene-based expected variation has a high predictive value of pathogenicity. We have tested this method in additional independent datasets with similar results. We are attempting to improve these predictions using additional variant and gene-level metadata and other variant-based in silico predictive techniques, such as PolyPhen and SIFT. These findings indicate that the background variation in asymptomatic individuals has high predictive value for variant pathogenicity. These data may be used to prioritize and filter observed variants for further review by researchers or clinical geneticists. This is one of the most urgent needs in clinical genomic interpretation [4] as a number of clinical labs [5-7] and direct-to-consumer groups [8] are providing WGS interpretation.


The Diversity Outbreed (DO) mouse population is a new heterogeneous stock mapping derived from the same eight inbred founder strains as the Collaborative Cross (CC) recombinant inbred strains. The DO mice have uniformly high levels of heterozygosity and genetic diversity, and thus provide a high-resolution mapping resource for identifying key genetic factors underlying complex traits and disease. Further, application of RNA-seq technology to the DO is more effective than the classical RNAseq methods because the DO mice are more diverse than the CC. This results in differences in the version of the reference assembly used by laboratories for analysis. Additionally, even when laboratories compute variation using the current reference assembly, they often wish to report variation on a RefSeqGene/LRG sequence or on a transcript or protein sequence. Using the UCSC LiftOver tool (http://genome.ucsc.edu/liftOver) users can map features from one coordinate system to another. We anticipate that this tool will be especially important with the upcoming reference assembly update (GRCh38).
1443W
Refining quality control for detection of rare structural variants with SNP arrays: presenting the ‘plumbCNV’ Package with an application to Type 1 Diabetes. N.J. Cooper, J.A. Todd. Medical Genetics, Diabetes and Inflammation Laboratory, Cambridge, UK.

Detection of rare copy number variants (CNVs) is increasingly seen as an important step towards understanding genetic influences on human disease. It has been shown that comprehensive quality control for Log-R-Ratio (LRR) intensity data is required to reduce the high false positive rate for rare CNV detection (Shirli et al., 2013). Such a pipeline should include examination of LRR distributions for problematic samples, followed by LRR-based principal component analysis (PCA) correction of batch effects. Once CNVs are called, additional tests should be applied to filter the set of CNVs detected. ‘plumbCNV’ is presented as an R package that automates these procedures. The code is open source, and scalable to deal efficiently with very large datasets exceeding system memory limits. Parameters are highly customisable and diagnostic plots and tables are produced throughout to allow the researcher to remain in control of the filters and thresholds being applied. This scripted pipeline provides a robust and necessary prerequisite for detection of rare CNVs using SNP arrays. Further assessment of CNV association with disease and phenotypes can therefore be readily and reliably conducted without using CGH or dedicated CNV arrays. The procedure is demonstrated with a comparison of rare CNV rates and characteristics between 6,292 Type 1 diabetes and 8,322 Healthy samples.

1444T
The Transition to Clinical NGS: How Well Do You Know Your Sequencing Pipeline? D. Corsmeier1, B. Kelly1, P. White1,2. 1) The Research Institute at Nationwide Children’s Hospital, Columbus, OH; 2) The Ohio State University, Columbus, OH.

As next generation sequencing technologies are rapidly adopted in clinical genetics settings, higher expectations must be placed on the bioinformatics pipelines used to transform hundreds of gigabytes of raw read data into the few lines of meaningful genetic variation that is useful to the clinician. Reducing these data by several orders of magnitude is a complex process with the capacity for error and inadvertent deviation from best practices. Notwithstanding these potential pitfalls, even if secondary analysis is performed correctly using accepted methods, parallelization and down-sampling techniques can introduce nondeterminism in the resulting variant call set. This brings into question the utility of a given computational approach in a clinical setting.

Churchill, our fully automated pipeline for secondary analysis, uses one of the most popular software combinations: the Burrows-Wheeler Aligner (BWA) for short read sequence alignment together with the Genome Analysis Toolkit (GATK) for variant calling and genotyping. Using novel parallelization approaches, we have overcome the computational bottleneck created by the exponential growth in the generation of genomic sequencing data, reducing analysis time from weeks to hours. In developing an algorithm so innovative in performance and efficiency, the validation of results is a necessity and the final hurdle towards potential clinical utility.

Committed to the repeatability of test results based on discrete digitized genomic data, we investigated three pipelines that use the BWA/GATK combination for secondary analysis: Churchill, GATK Queue, and HugeSEQ. Somewhat surprisingly, we discovered that nondeterminism can be introduced at virtually every step in the analysis if configuration parameters are not carefully selected. Of the analysis approaches tested, only Churchill preserved determinism while adhering to best practices, regardless of the level of parallelization. Further, we demonstrate that Churchill’s speed and precision do not come at the expense of quality of the output variant call set.

1445F
Overcoming genetic heterogeneity in rare Mendelian disease gene discovery: an improved network analysis approach. N. Dand1, F. Sprengel2, V. Ahlers2, M.A. Simpson1, R. Schulz1, R.J. Oakey1, T. Schlitt1,3,4. 1) Department of Medical and Molecular Genetics, King’s College London, United Kingdom; 2) Department of Computer Science, University of Applied Sciences and Arts Hannover, Germany; 3) Institute for Mathematical and Molecular Biomedicine, King’s College London; United Kingdom; 4) Novartis Institutes for BioMedical Research (NIBR), Basel, Switzerland.

In the past few years a number of groups have undertaken exome-sequencing studies and successfully identified genes responsible for rare Mendelian diseases, typically by sequencing a small number of unrelated patients and identifying shared genetic variants after filtering out those less likely to be causal. However, this strategy is not always successful and in particular can be limited by genetic heterogeneity, where a variant in any of several genes is sufficient to cause the same disease (for example due to shared functionality). We recently presented BioGranat-IG, a software tool which addresses this problem using gene or protein networks. BioGranat-IG identifies parsimonious, highly connected components in the network, containing variants (after filtering) for all (or most) patients. We hypothesised that these components correspond to likely disease-causing protein complexes and/or pathways. We demonstrated the validity of our method using simulated data for two rare Mendelian diseases known to have a genetically heterogeneous cause.

Since publication of this work, application to real disease data has highlighted three main areas where further improvement is possible: (i) large and highly polymorphic genes tend to be over-represented in proposed pathways; (ii) ‘hub’ genes, which are highly-connected in the network, also tend to be over-represented, and (iii) the correct filtering criteria for input gene lists are unclear. In the present work, we develop an entirely different algorithmic approach, using a score-based network prioritization method which removes the need for variant filtering, but retains the information from individual patients. Instead of simulating data, we use randomly rearranged exome sequence data generated at the KCL GISTT Biomedical Research Centre along with variant information obtained from the Online Mendelian Inheritance in Man (OMIM) database to generate realistic estimates of the performance of our method for different input data. We test the performance of BioGranat-IG, and propose that our new approach provides an efficient way to identify genetic mechanisms which, when disrupted, can result in disease.

1446W
SG-ADVISER tools: de-identification, identification and visual analytics of SG-ADVISER data. G. Erikson, P. Pham, W. Shipman, A. Torkamani. Scripps Translational Science Institute, La Jolla, CA.

Following variant annotation, a major task in genome interpretation is the identification and prioritization of relevant variants. Scripps Genome Annotation and Distributed Variant Interpretation Server (SG-ADVISER) is a web-server developed at The Scripps Translational Science Institute for the annotation of genetic variant data. To facilitate the interpretation of the SG-ADVISER output, we have developed two additional tools: SG-ADVISER UI and the SG-ADVISER Deidentification tool. The SG-ADVISER UI is a visualization tool that allows scientists with little or no programming experience to easily and quickly view, manipulate, sort, and filter the SG-ADVISER output file. A variety of custom and advanced filters allow filtering based on any combination of sample and annotation information. For example, it is possible to sort or filter on Coding Variants, Novel Variants, Cancer Genes or Chromosome Position etc. At any point during the process, the tool can provide summary statistics, and output the filtered results to a new file. The tool is written in Java and uses multi-threaded architecture and paginated view to make data analysis of large files (3GB+) possible on a desktop computer. To protect one’s privacy we created SG-ADVISER Deidentification, which is a tool that extracts the genotype(s) from the VCF files and implants clinically associated variants into the list of transmitted variants. For the clinically associated variants we used the Flagged SNPs 132 table from the UCSC Genome Browser. Identifier strips the clinically associated variants previously implanted and imports the genotype(s) back into SG-ADVISER annotated file. Both tools are cross-platform compatible and can be downloaded from the website: genomics.scripps.edu/ADVISER. This web portal can also be used for uploading de-identified data to the SG-ADVISER pipeline for annotation.
1447T Novel set of bioinformatics tools for performing distributed down-stream data analysis in very large GWAS and NGS projects. L. Eronen, T. Kaminen, P. Sevon. Bioinformatics platform, Biocomputing platforms Ltd, Helsinki, Finland.

The amount of data produced in large GWAS or NGS studies today can be measured in Terabytes. Complex data analyses, like imputing and down-stream analysis of imputed data as well as NGS data alignment and variant calling takes considerable amount of CPU time and may become a research bottleneck.

For many analysis tasks performance can be increased by data level parallelization; distributing data analyses by genomic regions, subjects and phenotypes. However, limited bandwidth between the database and external calculation resources, saturation of the disk system and task of dividing data set to smaller segments significantly limit the performance gains from distributed analysis. On the other hand, there’s often a lot of manual work involved in setting up a massively parallel run.

BCITOOLS is a software suite intended to reduce this overhead. Key components of the suite are a set of command line tools for submitting and managing distributed workflows, a workflow engine that can utilize different kinds of calculation resources (local servers, different calculation clusters or cloud), a set of file format converters that are applied automatically to make output of one analysis tool compatible with the next tool, and a set of split and merge tools for parallelizing the workload and merging results from subjobs. The system provides an API that enables easy integration of new tools.

To facilitate the massively parallel analysis of very large data sets we have developed a compressed virtual file system (BCFS), where massive data files are partitioned into tiles of manageable size, each covering only a subset of the genome, subjects and/or traits. All data files required during analyses are stored to BCFS, providing unified access to data from all calculation nodes. Physical storage may be distributed; e.g. imputation results can be stored as tiles on disks near the computation resources, while storing all clinical data on a local workstation.

In this presentation we evaluate feasibility and performance of BCITOOLS by performing a distributed analysis workflow of a dataset with 100,000 subjects, consisting of imputation and association analysis of the imputed data in a cloud environment.

1448F A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. H. Feng1,2, K.N. Connelly1, H. Wu1. 1) Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA 30322, U.S.A; 2) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, U.S.A.

DNA methylation is an important epigenetic modification that has essential roles in cellular processes including genome regulation, development, and disease, and is widely dysregulated in most types of cancer. Recent advances in sequencing technology have enabled the measurement of DNA methylation at single nucleotide resolution through methods such as whole-genome bisulfite sequencing and reduced representation bisulfite sequencing. In sequencing studies of DNA methylation, a key task is to identify methylation differences under distinct biological contexts, for example, in comparisons of tumor vs. normal tissue. A common challenge in these studies is that the number of biological replicates in each comparison group is limited due to the extreme costs of sequencing experiments. The small number of replicates leads to unreliable and unstable estimation of within-group variation, which can result in low power to detect differential methylation loci (DML). The variance shrinkage method has been widely applied to alleviate this type of problem since the microarray days. However, the nucleotide resolution DNA methylation data are discrete and are typically modeled by distributions where the variances are dependent on means. This poses some technical difficulties to variance shrinkage. Here we propose a novel statistical method to detect DML when comparing two treatments groups in studies of nucleotide resolution DNA methylation data. The sequencing counts data are described by a lognormal-beta-binomial hierarchical model, which provides a basis for inference at different levels of the data hierarchy, including different CpG sites. This model can be used to obtain empirical Bayes shrinkage estimates of the within-group variances, which are then plugged into a Wald test procedure for DML detection. We show through simulation that the proposed method provides greater power to detect DML compared to existing methods, particularly when the number of replicates is very low. The proposed method has been implemented in freely available Biocomputer package DSS.

1449W VAAST+ VSQR: making effective use of INDELS in disease-gene searches. S. Flygare1, B. Kennedy1, C. Huff2, M. Reese2, L.B. Jorde2. 1) Human Genetics, University of Utah, Salt Lake City, UT; 2) MD Anderson Cancer Center, University of Texas, Houston, Texas; 3) Omicia, Inc., Emeryville, California 94608, USA.

The Variant Annotation Analysis and Search Tool - VAAST - is a widely used tool for discovering disease-causing variants in personal genome data (Yandell et al., 2011). VAAST was used to discover a new human disease using next-generation sequencing data (Rope et al., 2011). Most recently, VAAST was used to identify the cause of Sturge-Weber Syndrome (Shirley et al., 2013). Despite its power, VAAST (and every search tool) is limited by the quality of its input data. Here we describe a new version of VAAST that addresses this problem. It works by employing VQSLOD scores provided by the GATK Variant Quality Score Recalibration procedure. We show that employing VQSLOD scores within VAAST results in a 30% increase in VAAST’s power to identify disease-causing mutations, particularly disease-causing indels. VAAST is thus a powerful search tool that can effectively use indels in disease-gene searches. We have also developed a second quality measure that extends the VQSLOD approach to allow identification of what is termed ‘induced false positives’. These induced false positives result from reads containing real, but rare, variants that drive the read’s alignment to an incorrect region of the genome, producing falsely induced variants with misleading high VQSLOD scores. Preliminary results indicate that these scores used alone improve VAAST’s power by 59%, compared to the 30% improvement seen with VQSLOD scores. Work is underway to integrate both quality scores into the VAAST algorithm, and even greater improvements are expected. As we will demonstrate, these new augmentations to VAAST make an effective tool for identification of disease-causing variants using even the noisiest indel calls.

1450T Sherlock: A Comprehensive Approach to Discovering Gene-Disease Associations in GWAS using eQTL. C. Fuller1, X. He2, H. Li3, L.B. Jorde1. 1) Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94113, USA; 2) Lane Center of Computational Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

The genetic mapping of complex disease has traditionally relied on the identification of variations that perturb either the function or expression of nearby genes. A strong body of evidence suggests that many distal (i.e. trans-acting) loci influence gene expression, but such associations are typically ignored in the standard gene-disease association paradigm. Moreover, large numbers of potentially informative associations from genome wide association studies (GWAS) are discarded when filtering for only genome-wide levels of significance. We recently introduced a Bayesian method and online service (Sherlock) that addresses both of these issues by matching association patterns between GWAS and expression quantitative trait loci (eQTL) to implicate genes that cannot be identified by GWAS alone. Here, we present an alternate statistical test that uses the empirical distribution of eQTL summary statistics to identify disease genes from GWAS data sets. This new approach permits a robust comparison of association patterns without assumed priors and in the presence of the systematic inflation that is common in real-world GWAS. We discuss the results from a matrix of tests spanning a dozen diseases and numerous eQTL tissue types. We identify candidate disease genes that replicate across different input GWAS and eQTL data sets. Importantly, we note instances of correlated regulation among trans-acting loci: the risk alleles for the set of SNPs that implicate a given gene all push its expression in the same direction. In isolation, GWAS tend to generate large numbers of low-to-moderate effect loci that identify nearby genes. A strong body of evidence suggests that many distal (i.e. trans-acting) loci influence gene expression, but such associations are typically ignored in the standard gene-disease association paradigm. Moreover, large numbers of potentially informative associations from genome wide association studies (GWAS) are discarded when filtering for only genome-wide levels of significance. We recently introduced a Bayesian method and online service (Sherlock) that addresses both of these issues by matching association patterns between GWAS and expression quantitative trait loci (eQTL) to implicate genes that cannot be identified by GWAS alone. Here, we present an alternate statistical test that uses the empirical distribution of eQTL summary statistics to identify disease genes from GWAS data sets. This new approach permits a robust comparison of association patterns without assumed priors and in the presence of the systematic inflation that is common in real-world GWAS. We discuss the results from a matrix of tests spanning a dozen diseases and numerous eQTL tissue types. We identify candidate disease genes that replicate across different input GWAS and eQTL data sets. Importantly, we note instances of correlated regulation among trans-acting loci: the risk alleles for the set of SNPs that implicate a given gene all push its expression in the same direction. In isolation, GWAS tend to generate large numbers of low-to-moderate effect loci, often poorly annotated and lacking any obvious functional consequence. Our method permits robust mapping of these loci to more functionally informative genetic regions, providing a genomewide mechanistic insight and statistical power over the GWAS results alone. The approach is broadly applicable and could easily incorporate other molecular associations, such as metabolites, non-coding RNAs, and epigenetic modifications. It represents a straightforward means of mining both current and future data repositories (e.g. the NIH Genotype-Tissue Expression initiative) to inform research that is starved for insight amidst a sea of data.
1451F Screening genome variants for disruption of regulatory activity with ZoomReG: capture, sequencing and computational identification of regulatory variants. T. Gasserfelder1, M.E. Edsall1, A.N. Dubinsky2, R. Chappel2, P. Ordoukhian3, S.R. Head3, The NEIGHBOR Consortium. 1) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA; 2) Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA; 3) Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, CA; 4) Department of Pediatrics, University of California San Diego, La Jolla, CA; 5) Next Generation Sequencing Core Facility, The Scripps Research Institute, La Jolla, CA.

Purpose: We identified and ranked genome variants likely to disrupt regulatory activity, including microRNA control of protein expression, RNA-binding protein (RBP) control of alternative splicing, and DNA-binding protein (DBP) influences. Background: Genome-wide sequencing identifies many variants that affect transcription regions (UTR), near intron/exon junctions and near gene structures. Efficient computational screens and low false positive rates are imperative for complete interpretation of the impact in individuals or disease cohorts. Methods and Results: Genome variants with potential to alter regulatory activity are screened through three computational modules comprising the ZoomReG process. ZoomReG was applied to 420 exomes sequenced with protocols optimized for reliable near-target capture. Collectively, over 7M chromosome sites were observed as variant in the 420 exomes with 93 pct present in 1000 Genomes. Results showed ZoomReG returned known regulatory variants and ranked them reliably for degree of disruption. ZoomReG applies its ZoomMiR module to determine which UTR and coding region variants are contained within potential seed sites for microRNA:mRNA binding. ZoomReG uses the prior HOMER algorithm and a database of >4000 RBP motifs to determine whether variants overlap RBP motifs. It uses HOMER to evaluate whether SNPs overlap DNA-binding motifs and ENCODE DNase-sensitive regions. ZoomReG focuses on 50 bases upstream and downstream of each variant site and returns a disruption score that represents whether binding centers on the site. E.g., a microRNA disruption prediction requires a SNP to appear in a satisfactory 'seed'. RBP binding is enhanced by tandem binding site motifs (e.g., YCAYCAY is stronger for Nova binding than YCA1). For RBP site disruption, SNPs in single motif matches have greater weight than in tandem matches. For a 3'UTR SNP shown in earlier work (NEIGHBOR GWAS of primary open angle glaucoma (POAG)) as associated with optic neuropathy, ZoomMiR in ZoomReG predicted a novel seed site disruption by the minor allele, validated with luciferase assay. Conclusion: ZoomReG provides critical screening functionality to identify genome variants that disrupt microRNA, RBP and DBP binding sites. Its ZoomMiR module revealed a molecular mechanism for a risk-associated SNP found in the NEIGHBOR POAG GWAS demonstrating this functionality is critical for complete genome sequence interpretation.


Current genome annotation pipelines frequently yield misleading results due to sequencing technology-specific biases, errors in the reference genome and other causes. Here we introduce a suite of tools for ‘multi-genome analysis’ that improve the identification and interpretation of genomic variants in the context of disease. The Family Genomics group (familygenomics.systemsbiology.net) at the Institute for Systems Biology has undertaken multiple collaborative projects related to understanding the genetic basis of disease, with special emphasis on neurodegeneration. We currently have high quality whole-genome sequence (WGS) data from 1,000 individuals, produced by Complete Genomics, and funded by the University of Luxembourg (www.uni.lu/lcsb). We analyze the data using custom workflows and the Ingenuity Variant Analysis platform (www.ingenity.com,variants). Our collective WGS dataset serves as a superb resource for modeling systematic failures and biases in the technology. Use of multi-genome models improves our ability to analyze each individual genome, leading to fewer false positive and false negative findings. Simultaneous coverage analysis of several hundred genomes enables detailed normalization of the coverage profiles of individual genomes. This enables precise analysis of CNVs and the identification of large deletions that were previously undetectable. Some of the deletions we discovered explain the observed pattern of disease inheritance in the families we are studying. When analyzing personal genomes, certain genes frequently show up as mutated. We identified a 39Kbp deletion of chromosome 22q13.12 associated with 39 different rare variants. Fine-grained annotation of these problematic regions leads to fewer spurious findings – a particular concern in clinical settings. We compiled genome-wide maps of regions prone to accumulating ‘no call’ failures and informed the alignment pipeline to display these regions with increased heterozygosity, and to being identified as ‘identical by descent’ among unrelated individuals. We make available several resources for improving the quality of personal genome analyses, individually and in the context of family pedigrees. Some of the resultant improvements to sensitivity and specificity are crucial for achieving clinical-grade genome interpretation.

1452W W4CSEQ – a web server to process enzyme-based and sonication-based 4C-seq data. F. Gao1, 2, W. Lu1, K. Wang1. 1) Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA 90089, USA; 2) Zilkha Neurogenetic Institute, Department of Psychiatry and Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA.

Circular chromosome conformation capture, when coupled with next-generation sequencing (4C-Seq), can be used to identify genome-wide interaction of a given locus with all of its interacting partners. Both restriction enzyme digestion and sonication methods can be used to fragment cross-linked chromatin for proximity-based ligation. We recently applied 4C-Seq to characterize the interactome of an enhancer element of Pou5f1 gene in pluripotent stem cells. We compared replicate consistency as well as similarity of 4C-Seq data generated by both enzyme-based and sonication-based fragmentation methods. In general, we found good correlation (>0.6) for inter-chromosomal interactions identified by both methods when comparing biological replicate data. However, sonication generated less distal intra-chromosomal interactions compared to enzyme-based method. As noted, data generated by both methods showed correlation with DNA early replication timing and enrichment of active histone marks in the enhancer interactomes. In addition, the interacting sites identified from sonication-based data are enriched with ChiP-Seq signals of transcription factors Oct4, Klf4, Esrb, Tcfcp21 and Zfx that are key regulators for reprogramming and pluripotency. We found both methods are valuable tools for exploring long-range chromosomal interactions. To help bench scientists with minimal NGS expertise to utilize 4C-Seq technique in the research, we have integrated our bioinformatics pipelines into a web-based server that can automatically process 4C-Seq raw data (FASTQ format) generated by a high-throughput Illumina sequencer, such as HiSeq 2000. This web server (w4cseq.usc.edu) automatically takes user specified input file(s), performs reads mapping to the selected reference genome (currently only human and mouse genomes are included), calculates statistically significant interacting genomic regions, and generates relevant plots for illustration. For an input FASTQ file containing 10 million reads, the user is expected to receive the result in one day.
PhaseLift: a novel procedure to save time in imputing study genotypes.

PhaseLift: a novel procedure to save time in imputing study genotypes. M.M. Gorskii, T.W. Winkler, K. Stark*, M. Muller-Nurasyid, J.S. Redd, B.H. Weber, P. Heid*. 1) Department of Epidemiology and Preventive Medicine, Medical University Center Regensburg, Regensburg, Bayern, Germany; Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany; 2) Department of Internal Medicine II, University Medical Center Regensburg, Regensburg, Germany; 3) Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; 4) Department of Medicine I, University Hospital Großhadern, Ludwig-Maximilians University, Munich, Germany; 5) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology and Chair of Genetic Epidemiology, Ludwig-Maximilians-University of Munich, Munich, Germany; 6) Institute of Human Genetics, University of Regensburg, Regensburg, Germany.

Genome-wide association (GWAs) studies usually apply imputation techniques to complement genome-wide SNP chip genotypes. The recent genotyped imputation methods separate phasing from imputing. This allows for updating the imputation when a new reference panel is released without repeating the phasing step. When the study data is on an older genome build than the reference data, the current lift-over procedure is to harmonize on the genotype-level (pre-phasing lift-over) with consecutive re-phasing and re-imputing which does not fully utilize the advantage of the current two-step imputation method. We propose a novel lift-over procedure that phases genotypes on the original build and harmonizes the study haplotypes with the reference data (post-phasing lift-over; implemented in the software PhaseLift). This avoids re-phasing of study data when switching to a new reference panel. We use the KORA study (1,644 unrelated individuals) to evaluate the lift-over procedures and compare the estimated allele dosages with the additionally typed variants from the Illumina Cardio-Metabochip. To contrast the performance of both procedures, we imputed the GWA SNPs with both procedures and compare the estimated allele dosages with the additionally typed genotypes. We found that both approaches perform equally well with mean concordances of 93% and 93.4%. To see if one procedure yields higher imputation qualities (RSQs), we computed the difference of RSQs imputed by both procedures and found little difference (mean difference = 0.007, median difference = 0.007, 95% confidence interval: -0.113, 0.127, p-value = 0.90). To examine whether the advantage depends on the quality of the reference panel, we compared SNPs imputed by both procedures, categorized by RSQ. We found that a comparable number of SNPs is poorly (RSQ < 0.3: 78.75%, 78.78%), medium well (0.3 < RSQ < 0.8: 9.99%, 9.97%) and well imputed (RSQ > 0.8: 11.26%, 11.25%) by both procedures, respectively. By using our new procedure you can save nearly 2 months of parallel computing on an 8 core cluster, when re-imputing 3 times with the 1000G GIANT ALL reference panel. We demonstrate that our proposed post-phasing lift-over approach might encourage study partners to quickly accommodate updated reference builds to improve the information content of their data.

CIDRSeqSuite 4.0: A Toolbox for Next-Generation Sequencing Workflow Development.


The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and access to genetics consultation to investigators working to discover genes that contribute to disease. CIDRSeqSuite is a set of software tools that has been used to aid in next-generation sequencing analysis at CIDR since 2009. A core feature of CIDRSeqSuite is a custom developed analysis pipeline that is designed to keep pace with changes to the predominant sequencing analysis tools used by the community, and similarly the choice of underlying technologies used in the implementation of these tools has evolved to keep pace with the larger number of analyses that must be done in parallel and in a fault-tolerant fashion. The initial workflow implemented for CIDRSeqSuite was designed to run on a single large server; samples were processed a few at a time in a single process. The current version of CIDRSeqSuite, however, focuses on individual tasks and the interdependencies among them. When an analysis is submitted to the CIDRSeqSuite server, all required tasks and their interdependencies are stored in a relational database. Such tasks include alignment, variant annotation, and even the demultiplexing of an individual sample from a flowcell. Tasks are submitted by the server process to an SGE-enabled cluster of over 25 nodes. As each task finishes, its status is stored in the database; any tasks whose dependencies are complete are then submitted themselves. Should a task fail to complete, it is immediately resubmitted to the compute cluster; if the number of allowable retries for a given task is exceeded, the task is assigned to a lab or informatics personnel to aid in troubleshooting. Notifications are also sent when analyses finish so that the data generated can be reviewed. Tasks can be submitted to the compute cluster via command line tools or a graphical user interface. Advantages to this distributed task-based approach include considerably faster demultiplexing: processing time is reduced from 6-8 hours to 30 minutes per flowcell. In addition to providing the apparatus for workflow development and execution, CIDRSeqSuite also encompasses several other tools including the generation of various QC reports, concordance reports (both between sets of sequencing data and between sequencing and genotyping data), and stand-alone paired-end and single-end demultiplexers.

Aneuploidy and normal cell contamination aware approach to detect copy number variations in cancer using next generation sequencing data.

Aneuploidy and normal cell contamination aware approach to detect copy number variations in cancer using next generation sequencing data. R. Gupta, S. Katragadda, D. Vyavahare, K. Sandhu, V. Veeramachaneni, R. Hanharan. Strand Life Sciences, 5th Floor, Kirloskar Business Park, Bellary Road, Hebbal Bangalore - 560024, Karnataka, India.

Background and Objectives: Recent growth in next generation sequencing (NGS) data has enabled us to detect copy number variation (CNV) at an unprecedented resolution. The objective of this study is to develop an approach that can: 1) Identify the CNV regions in the cancer genome and assign absolute copy number (CN); and 2) Compare CNV regions from different patients to identify regions that are commonly amplified or deleted, thereby highlighting genes implicated in cancer.

Challenges: Several technical and biological challenges inhibit the discovery of true segments and assignment of absolute CNs. In particular, biological challenges include 1) Aneuploidy of cancer cells but many approaches assume diploid genome; 2) Contamination by normal cells but many approaches assume diploid genome; 3) Heterogeneity in tumour cells i.e. there may be polyclonal tumours with different CNVs in each clone; and 4) Normal cell contamination. Methods: Most of the approaches for detecting CNVs from NGS data are based on 1) read depth; 2) distance orientation of read pairs; and 3) split reads. We used a method based on read depth and first compute the log-ratio of read depth in cancer and normal samples for fixed length windows, followed by Wavelet transformation of ratios to reduce the effect of random noise. An EM algorithm based probabilistic Gaussian mixture model is then built to model different CN states, and biological parameters of the sample, average ploidy and % normal cell contamination, are estimated. Finally, we used two segmentation approaches on the ploidy and contamination corrected log-ratio to obtain segments and corresponding CNVs. First is naive and heuristic approach, which quickly identifies gain/loss regions without qualifying the degree of gain or loss; and 2) popular CBS approach, which can distinguish different gain (or loss) regions. This CNV detection approach is integrated in Avadis NGS, which is our software tool for the processing and comprehensive end-to-end analysis of NGS data. Experiments and Results: We demonstrated the efficacy of the CNV detection approach on both simulated data and publicly available real sequencing data. For simulation set, we simulated log-ratio data to cover different scenarios by varying sample ploidy, % of normal cell contamination, number of CN states, % of data noise, etc. We also used publicly available sequencing data of cancer cell lines and tumor samples from NCBI SRA and construct CN profiles for multiple cancer types.

Advantages to this distributed task-based approach include considerably faster demultiplexing: processing time is reduced from 6-8 hours to 30 minutes per flowcell. In addition to providing the apparatus for workflow development and execution, CIDRSeqSuite also encompasses several other tools including the generation of various QC reports, concordance reports (both between sets of sequencing data and between sequencing and genotyping data), and stand-alone paired-end and single-end demultiplexers.
1457F

Killer-cell immunoglobulin-like receptors (KIRs) are one of the most polymorphic genes, and besides their sequence diversity there also a high homology among these receptors. This makes KIR typing exceptionally difficult after the sequencing. The even more complex since not all KIR genes are present in an individual and one gene can be present in many copies. As next generation sequencing have spread to most of molecular genetic laboratories, there is a need for a method for KIR typing based on short reads. We are presenting an algorithm to determine the KIR type from NGS samples. The method is capable to predict the copy number of genes and we are also presenting quality check measures to estimate the goodness of typing.

1458W
Estimating exome genotyping accuracy by comparing to data from large scale sequencing projects. V. Heinrich1, T. Kamphans2, J. Stange2, D. Parkhomchuk3, J. Hoche4, T. Dickhaus5, P.N. Robinson6, P.M. Krawitz7, 1) Institute for Medical Genetics and Human Genetics, Charité Berlin, Germany; 2) Max Planck Institute for Molecular Genetics, Berlin, Germany; 3) Department of Mathematics, Humboldt University Berlin, Germany; 4) BCRT, Berlin, Germany; 5) Department of Medical Informatics, University of Louvain, Belgium.

Next-generation sequencing (NGS) based methods, such as exome analysis, are currently used as a tool for diagnostics. This increases the need for platform-independent methods of quality control. To date, quality scores have been used on a base- or genotype-level type to indicate the reliability of single base or variant calls. However, the great variety of NGS platforms and analysis pipelines hinders the direct comparison of genotype-specific quality scores between platforms or for entire data sets and currently no criteria exist for assessing the overall quality of an exome. We present a genotype-weighted metric to compare all exome variants identified in a single sample together to an appropriate high-quality reference data set, with which we estimate the exome-wide genotyping accuracy based simply on the reported variants and without any further knowledge about the data generation. Our method represents a new way to evaluate the quality of entire whole-exome sequencing data in addition to current recommendations for sequencing depth and genotype likelihoods. The distance value of our metric corresponds to a quality parameter for an entire exome and allows comparing the quality of multiple exome datasets from the same or different NGS platforms. Based on simulated accuracy groups for variant calls we were able to assess the quality of an exome sample without detailed knowledge about the applied enrichment and sequencing technology or about the bioinformatics pipeline that was used to align the reads and call the genotypes. We envision that our approach to estimate the genotyping accuracy of exomes will facilitate the quality assessment of NGS data.

1459T
Highlander: variant filtering made easier. R. Holasen, M. Vikkula. Laboratory of Human Molecular Genetics, de Duve Institute (Université catholique de Louvain), Brussels, Belgium.

The field of human genetics is being revolutionized by exome and genome sequencing. A massive amount of data is being produced at ever-increasing rates. The data is currently introduced as a tool for mutation detection into routine diagnostics. This increases the need for platform-independent methods of quality control. To date, quality scores have been used on a base- or genotype-level type to indicate the reliability of single base or variant calls. However, the great variety of NGS platforms and analysis pipelines hinders the direct comparison of genotype-specific quality scores between platforms or for entire data sets and currently no criteria exist for assessing the overall quality of an exome. We present a genotype-weighted metric to compare all exome variants identified in a single sample together to an appropriate high-quality reference data set, with which we estimate the exome-wide genotyping accuracy based simply on the reported variants and without any further knowledge about the data generation. Our method represents a new way to evaluate the quality of entire whole-exome sequencing data in addition to current recommendations for sequencing depth and genotype likelihoods. The distance value of our metric corresponds to a quality parameter for an entire exome and allows comparing the quality of multiple exome datasets from the same or different NGS platforms. Based on simulated accuracy groups for variant calls we were able to assess the quality of an exome sample without detailed knowledge about the applied enrichment and sequencing technology or about the bioinformatics pipeline that was used to align the reads and call the genotypes. We envision that our approach to estimate the genotyping accuracy of exomes will facilitate the quality assessment of NGS data.

1460F
Comparison of Unified Genotyper and SAMTools as variant callers across 84 exomes. K. Hatrick1, H. Ling, E. Pugh, S. Griffith, B. Craig, B. Marcyos2, S. Branch-Deneny3, K. DeMey4. 1) Center for Inherited Disease Research (CIDR), Johns Hopkins Univ, Baltimore, MD; 2) UC-Berkeley, CA; 3) Omixon, Omixon Biocomputing, Budapest, Hungary.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR has been utilizing small tools (v0.1.14) for perform variant calling on exomes (Barnhart, 1693/Poster, ASHG 2011). Calls were made and filtered per individual and then all unique calls within a study were back-genotyped for all samples from the samtools bcftools output to create a multisample vcf for SNVs. We evaluated the Genome Analysis Toolkit 2 (GATK), to see if the significant advances in the GATK 2 framework (namely the addition of recalibrating indels with Base Call Quality Score Recalibration (BQSR) and creation of Reduced Reads Bams (RR) for analysis) improved variant calls. Briefly, 84 exomes (Agilent® SureSelect™ XT Human All Exon v4, Illumina® HiSeq™ 2000), were aligned with BWA-0.5.10 using the 1000 genomes phase 2 reference genome, duplicate molecules flagged with Picard (v1.74), Local Realignment, BQSR (GATK 2.1-9, 2.1-9) and RR (GATK 2.1-9) to create analysis BAM files. Multi-sample calling and filtering was done with Unified Genotyper (UG) and Variant Quality Score Recalibration (VQSR) (GATK 2.1-9). These calls were then compared to calls made from our existing samtools pipeline. The mean per sample on exon Ti/Tv for known SNVs (dbSNP 129) for both sets was ~3.2. For novel snps, the Ti/Tv per sample mean was 2.8 for the UG+VQSR calls and 2.1 for the samtools calls. The mean per sample count of exon SNVs was 20,521 (GATK) and 20,628 (samtools). Across 9 trios in the sample set, the on exon per variant site mendel error rate was 3.15% (GATK), 2.79% (samtools). Ti/Tv variability concordance rate between samtools and UG+VQSR was 90.77%, 92.97%, 90.97%, 92.99%, 90.94%, 98.10%, respectively. We have phased in a new workflow utilizing the GATK 2 framework which incorporates multi-sample calling/filtering with UG and VQSR and continue to evaluate all components to the NGS analysis pipeline to provide high-quality variants calls for NGS applications.

1461F
postMUT: A Statistical Tool for Combining Predictions of Missense Mutation Functionality using Capture-Recapture Methods. S. Hicks1,2, S.P. Plott1, M. Kimmel1,2, S. Paddison1. 1) Department of Statistics, Rice University, Houston, TX; 2) Department of Pediatrics and Human and Molecular Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Bioengineering, Rice University, Houston, TX.

Computational or in silico methods such as SIFT or PolyPhen-2 have been widely used to predict the impact of missense mutations on protein function, but these methods often report conflicting results. This leaves researchers without guidance in how to prioritize the mutations identified for further evaluation. Ad hoc combinations of in silico methods (e.g., at least tens of thousands of variants requires annotation drawn from various sources, as well as advanced filtering capabilities. We have developed Highlander, a Java software coupled to a MySQL database, in order to centralize all variant data and annotations from the lab, and to provide powerful filtering tools that are easily accessible to the biologist. Data can be generated by any NGS machine (such as Life Technologies’ SOLID or Ion Torrent, or Illumina’s HiSeq) and most variant callers (such as Life Technologies’ LifeScope or Broad Institute’s GATK). Variant calls are annotated using DBNSFP and SnpEff, then imported into the database. The Highlander GUI easily allows for complex queries to this database, using shortcuts for certain standard criteria such as ‘sample-specific variants’, ‘variants common to specific samples’ or ‘combined-heterozygous genes’. Users can then browse through query results using sorting, masking and highlighting of information.

1461W
postMUT: A Statistical Tool for Combining Predictions of Missense Mutation Functionality using Capture-Recapture Methods. S. Hicks1, S.P. Plott1, M. Kimmel1,2, S. Paddison1. 1) Department of Statistics, Rice University, Houston, TX; 2) Departments of Pediatrics and Human and Molecular Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Bioengineering, Rice University, Houston, TX.

Missense mutations are among the most likely to contribute to disease, yet prioritizing these specific deleterious mutations is a difficult task due to the large number of potential deleterious mutations that must be evaluated. This can be particularly challenging for individuals where multiple biologically important mutations may be present in an individual and one gene can be present in many copies. As next generation sequencing have spread to most of molecular genetic laboratories, there is a need for a method for KIR typing based on short reads. We are presenting an algorithm to determine the KIR type from NGS samples. The method is capable to predict the copy number of genes and we are also presenting quality check measures to estimate the goodness of typing.
1462T Improved access to data sets via metadata-driven searches of experimental conditions at the new ENCODE Portal. E.L. Hong1, B.C. Hitz2, E.T. Chan1, D.T. Erickson3, N.R. Peddutin1, G. Roe4, K. Rosenblom5, L.D. Rowe6, C.A. Sloan7, J.S. Stratton7, G. Barber7, G.A. Binkley8, J. Garcia9, D. Karolchik2, K. Learned7, B. Lee2, S. Miyasato1, G. Moro2, M. Simson1, E. Weiler7, W.J. Kent1, J.M. Cherry7. 1) Department of Genetics, School of Medicine, Stanford University, Stanford, CA 94305; 2) Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA, 95064.

The Encyclopedia of DNA Elements (ENCODE) Project is a collaborative project to create a comprehensive catalog of functional elements in the human and mouse genomes. Now in its 8th year, the ENCODE Project has grown to include additional experimental techniques and genomic elements to survey. All experimental data (using more than 40 different experimental techniques to survey DNA-binding proteins, RNA-binding proteins, the transcriptional landscape, and chromatin structure in 400+ cell lines and tissues) and computational analyses of these data are submitted to the Data Coordination Center (DCC) for validation, tracking, storage, and distribution to community resources and the scientific community. Metadata describing important experimental conditions, such as the biological samples, specific reagents, and protocols necessary to replicate the assay, have been expanded and are being submitted to a newly-formed DCC. As the volume of data increases, the identification and organization of data sets becomes challenging. Here, we describe the design principles of how metadata are organized and annotated at the ENCODE DCC in order to facilitate the identification and comparison of data sets generated by the ENCODE project. The organization of the metadata will allow flexible and powerful searches on the revamped ENCODE Portal, the public website of the ENCODE Project, as well as support intuitive displays of biological significance, and protocols used for an experiment. Data from the ENCODE project can be accessed via the ENCODE portal (http://www.encodeproject.org) and the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

1463F Network analysis on established schizophrenia loci reveals significant genetic overlap with autoimmune and cardiometabolic disease. L.M. Huckins, E. Zeggini. Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The investigation of shared genetic factors underpinning diverse complex traits can aid the identification of common aetiopathology and give insights into shared biological pathways. We have developed a network-based method which searches for potential links between genes based on published GWAS co-citation. In order to investigate a set of association signals, we create a network in which each node represents a gene. Genes are selected on the basis of the position of the index associated variant, for example all genes within a recombination interval or genes residing either side of the signal of interest. We then create an edge between all pairs of genes which have previously been co-cited in a GWAS of any trait, or which have been cited in two separate GWAS for the same trait. We group genes into three categories; those which had been cited with a p-value lower than 10^{-4}, 10^{-6} and 10^{-8}. This type of analysis is hypothesis-free, using only previous GWAS to define edges between genes. Our initial analysis focused on the data base of GWAS associations reported in [1]. We tested the network analysis method on a set of 14 established schizophrenia (SCZ) loci. We computed clustering coefficients and compared these to the coefficients found when generating a million random gene sets of the same size. We found that the SCZ gene set was significantly enriched for genes which had been co-cited in GWAS of other traits at all three significance thresholds (p=0.00015 at the 10^{-4} association threshold). We observed overlap (4/14 genes) between GWAS of autoimmune diseases (Crohn’s disease, rheumatoid arthritis, multiple sclerosis, amyotrophic lateral sclerosis, obesity, diabetes) and SCZ. In addition, we also observed overlap between cardiovascular-related disease GWAS loci (type 2 diabetes, glucose levels, coronary artery disease and hypertension) and SCZ (5/14 genes). Both of these findings reflect well-established epidemiological links; studies have shown a possible autoimmune etiology for schizophrenia and non-affective psychosis [2], as well as indicating a possible shared genetic basis for schizophrenia and T2D[3]. We are extending our analyses to the full NHGRI catalog of published GWAS. [1] AD Johnson, CJ O’Donnell (2009) BMC Med Genet 10:8 [2] WW Eaton et al (2010) Bioinformatics 26(7): Bio384-90 [3] P.I LIN, A.R. SHULDINER (2010) Schizophrenia res 123, no2-3, pp. 234-243.

1464W SNP discovery in family sequencing datasets using Bayesian networks. A. Iruday, G. Martí, Biology, Boston College, Chestnut Hill, MA.

Genome-wide association studies are powerful tools to detect SNP variants that segregate at intermediate frequency with modest effect for complex traits. There has been increased interest in discovering rare alleles of large effect to see if these variants can explain the heritability of complex traits. The rare alleles are potentially enriched in families with multiple affected individuals there has been increased interest in family based sequencing studies. Yet, most variant discovery methods do not explicitly model family relatedness when analyzing data. Here we describe a variant discovery approach, which models the pedigree relationships and sequencing data as a Bayesian Network (BN). A BN represents the dependencies of the data as a directed acyclic graph. The application of a belief propagation algorithm to compute posterior genotype probabilities on a simulated trio dataset shows that MAF 0.01-0.004 and genotype discrepancies (0–15%) vary as a function of average read coverage (20x-5x), with the offspring sensitivity and accuracy performing the best. Further work is actively being pursued to improve the performance metrics of Pgmap and to apply it to different pedigree structures as well as empirical datasets.

1465T Mapping Diagnostic Test Requisition Data to the Human PhenoType Ontology. R. James, C. Shaw, Molecular and Human genetics, Baylor College of Medicine, Houston, TX, USA.

The clinical implementation of sequencing diagnostics encompassing coding regions (i.e., whole exome) has great potential to contribute to the diagnosis of human disease. The exome approach is made difficult because of the large number of variants observed in individual patients. Efficient use of available clinical information on patients may improve the curation and prioritization of observed variants. Unfortunately, phenotypic clinical inputs for genetic testing are rarely complete and often do not strictly adhere to a controlled vocabulary. We hypothesized that natural language sentences applied to clinical notes, indication forms and/or elements of medical records can standardize and improve indication data. We used these techniques in conjunction with hierarchical semantic relationships of The Human PhenoType Ontology to develop methods for processing indication data. Improved similarity metrics were developed and systematically evaluated via their performance in Monte Carlo simulations of possible feature combinations that may occur in patients being sent for genome wide diagnostics. These methods can be used to analyze the correspondence between the requisition content and the ultimate diagnosis obtained. We conclude that such approaches are useful to improve the speed and accuracy of exome analysis.

1466F Multiple Testing Correction method for Linear Mixed Model. J.W. JOO1, E. Koatem2, E. Kang2, B. Han2,4, E. Esk0n2,3, 1) Bioinformatics PhD program, University of California, Los Angeles, Los Angeles, CA, USA; 2) Computer Science, University of California, Los Angeles, Los Angeles, CA, USA; 3) Division of Genetics, Brigham & Women’s Hospital, Harvard Medical School, Boston, MA, USA; 4) Department of Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA.

Multiple hypothesis testing is a major issue in genome-wide association studies which often analyze millions of SNPs. There are several methods to estimate a per-marker p-value threshold in order to obtain a significant result given multiple markers. The permutation test is widely believed the gold standard for multiple testing correction for traditional association tests and accurately takes into account the correlation structure of the genome unlike the traditional Bonferroni correction. Since the permutation test is computationally very expensive, several alternative methods have been presented to speed up the permutation test. Recently, mixed model association has become a popular approach for genome wide association studies as it can correct for population structure. Population structure complicates association analysis by inducing spurious correlation between genotypes and phenotypes and may cause false positive associations. Unfortunately, permutation is not applicable for mixed models because if we either permute phenotypes or genotypes, the true correlation structure between individuals is eliminated and leads to an inflation of p-values. Theoretically, bootstrapping can be applied to mixed model, however, it is often computationally impractical for large datasets. In this paper, we propose an efficient and accurate multiple testing correction method for linear mixed model, slideLMM. The key idea behind slideLMM is that we utilize the kinship matrix to transform the correlation structure between the variants and then take advantage that the statistics will only follow a multivariate normal distribution. Utilizing this approach, slideLMM corrects for the population structure to give an accurate per-marker threshold even the individuals have different degrees of relatedness. Applied to the Hybrid Mouse Diversity Panel (HMDP) which is a mouse association study panel with significant amount of population structure, our method shows better performance in both speed and accuracy compared to previous methods.

The identification of disease-causing mutations in next-generation sequencing (NGS) data requires efficient filtering techniques. In patients with rare recessive diseases compound heterozygosity of pathogenic mutations is the likeliest inheritance model if the parents are non-consanguineous. We developed a web-based compound heterozygous filter that is suited for data from NGS projects and that is easy to use for non-bioinformaticians.

We analyzed the power of compound heterozygous mutation filtering by deriving background distributions for healthy individuals from different ethnicities and studied the effectiveness in trios as well as more complex pedigree structures. While usually more than 30 genes harbor potential compound heterozygotes in single exomes, this number can be markedly reduced with every additional member of the pedigree that is included in the analysis. In a real data set with exomes of four family members, two sisters affected by Mabry syndrome and their healthy parents, the disease-causing gene PIGO, which harbors the pathogenic compound heterozygous variants, could readily be identified. Compound heterozygous filtering is an efficient means to reduce the number of candidate mutations in studies aiming at identifying recessive disease genes in non-consanguineous families. A web-server is provided to make this filtering strategy available at www.gene-talk.de.


As sequencing technologies continue to evolve and the use of sequencing data makes its way from research into the clinic and hospital, the proliferation of data will continue to accelerate. With this trend and the application of this data to personalized medicine, new challenges in data storage, sharing, security, analysis and retrieval of information will arise. While many of these issues are only now starting to be addressed and anticipated, there is a considerable dearth of readily available solutions to these problems. The creation of data repositories capable of managing genomic information in a manner that enables streamlined access to data has emerged as a critical requirement as the application and use of such data progresses. One highly relevant example of a data repository solution that fulfills multiple needs for a variety of different users is The Cancer Genomics Hub (CGHub). CGHub is a genomic data repository built to support all three major NCI cancer genome sequencing programs: TCGA, TARGET, and the CGCI. CGHub was launched in 2012, hosted by UC Santa Cruz and with only TCGA data online, has more than 44,000 data files totaling more than 500 Terabytes with capacity to grow quickly to 5 Petabytes as additional datasets become available. CGHub is co-located with a biocompute farm that enables cancer researchers the ability to seamlessly access the data files for subsequent analysis using a variety of commercially and/or freely available tools. This repository was built with products and technologies developed at Annai Systems. In the CGHub example, data are stored in a vast public repository enabling widespread access to a large number of researchers and clinicians. There are also a growing number of smaller sized private repositories used to inform drug discovery, disease diagnosis and patient treatment. While some of the requirements of these repositories are quite similar to one another, there are a number of differences with respect to how the data are used, who will access it, and what type of regulatory and security considerations must be adhered to. Because of these differences, having a set of tools that can be used to provide flexible, scalable solutions that can address multiple use cases is of paramount importance. We will discuss how Annai Systems’ portfolio of products and services can be used to create ‘big data’ repositories that can facilitate access to, and sharing of, sequence data and related meta-data.

Churchill: A Cloud-Enabled, Ultra-Fast Computational Approach for the Discovery of Human Genetic Variation. B. Kelly, J. Fitch, D. Cors-mann, D. Knowles. 1) Research Institute at Nationwide Children’s Hospital, Columbus, OH; 2) The Ohio State University, Columbus, OH.

Next generation sequencing (NGS) has revolutionized genetic research, empowering dramatic increases in the discovery of new functional variants. The technology has been widely adopted by the research community and is now seeing rapid adoption clinically, driven by recognition of NGS’s diagnostic utility and enhancements in quality and speed of data acquisition. Compounded by declining sequencing costs, this exponential growth in data generation has created a computational bottleneck. Current analysis approaches can take weeks to complete resulting in bioinformatics over-heads that exceed raw sequencing costs and represent a significant limitation for those utilizing the technology. Churchill is a computational approach that overcomes these challenges, fully automating the analytical process required to take raw sequencing instrument output through the complex and computationally intensive processes of alignment, post-alignment processing, local realignment, recalibration and variant calling. Through implementation of novel parallelization techniques we have dramatically reduced the analysis time for whole human genome resequencing from weeks to hours, without the need for specialized analysis equipment or supercomputers. As increasing numbers of molecular diagnostic laboratories implement NGS in clinical settings, Churchill provides a solution to the data analysis challenges these laboratories will immediately face. Compared with alternative analysis pipelines, Churchill is simpler, faster, deterministic and capable of running on all popular Linux environments. Furthermore, Churchill optimizes utilization of available compute resources and scales linearly in a cost-effective manner, enabling complete human genome resequencing analysis in ten hours with a single server, three hours with our in-house cluster and under two hours using a larger HPC cluster. Churchill is cloud-compatible and we demonstrate the extensive degree of parallelization Churchill can achieve using Amazon’s Elastic Cloud Compute (EC2) instances. Not only does this allow laboratories to potentially reduce analysis time by leveraging the cloud's ability to easily scale, but it also enables low-cost resequencing analysis without the need to invest in expensive and hard to find high-performance computing clusters. Churchill eliminates the NGS bioinformatics overhead and is a prime candidate to overcome the bottleneck even faster sequencing will create.


Despite advancements in analysis methods, considerable human genetic variation remains inaccessible to short read sequencing. We combine targeted assembly and alignment of short reads with coverage- and base quality-enabled genotyping on the Amazon Elastic Compute Cloud to reanalyze exome data from the 1000 Genomes Project, benchmarking our results against Sanger sequences for a subset of samples and disease associated genes. Our results demonstrate high detection accuracy for variants, including insertions and deletions, up to 99.7% sensitivity with 1.7% false discovery. Applying our method broadly revealed significant population-dependent allelic diversity in regions previously considered inaccessible to variant detection. Specifically, microsatellite polymorphisms and pathogenic alleles in genes associated with Niemann-Pick disease, maple syrup urine disease, Usher syndrome, cystic fibrosis, and heritable amyotrophic lateral sclerosis. Scalable computational resources and highly accurate analyses, such as those described here, should accelerate clinical adoption of short read sequencing technologies by improving the power to detect variants.
1471T

Genomics in Clinical Research on the DNASTAR Cloud. M. Keyser1, J. Carville1, T. Schwei1, T. Derue1 PhD, A. Pollack-Berti PhD1, D. Nash1, J. Stierer1, S. Baldwin1, R. Nelson PhD1, K. Dullea1, J. Schroeder1, P. Pinkas PhD1. G. Plunkett III PhD2, F. Blattner PhD2,3, 1) DNASTAR, Inc., Madison, WI; 2) University of Wisconsin, Department of Genetics, Madison, Wisconsin, USA; 3) Scarab Genomics LLC, Madison, Wisconsin, USA.

DNASTAR offers an integrated suite of software for assembling and analyzing sequence data from all major next-generation sequencing platforms supporting key workflows on both a desktop computer as well as the DNASTAR Cloud. The cancer genomics workflow that integrates some of the most powerful functionality in the software includes assembling and analyzing multiple samples using one reference template; probabilistic identification of SNPs, small indels and genotype calls with known variants correlated to their dbSNP and COSMIC IDs and GERP reference data; review and filtering of SNPs from multiple samples within a single project; identification of structural variations; and, for large multiple-sample projects with hundreds of individual data sets, tools for SNP quantitation, filtering, set comparison, clustering and indication of the gene disruption impact from called SNPs. In addition, DNASTAR offers multi-sample copy number variation reporting for further analysis. Interactive views within the software facilitate fast, comprehensive analysis, helping scientists move quickly from raw next-gen sequencing data to genetic and genomic impact, including gene ontology. By using innovative algorithms within the software, scientists can have all of the assembly and analysis capabilities available to them on either their desktop computer or the DNASTAR Cloud, supporting large data sets generated by any or all of the next-gen sequencing instruments and platforms.

1472F

A general framework for estimating the relative pathogenicity of human genetic variants. M. Kircher1, D.M. Witten2, G.M. Cooper1, J. Shendure1. 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Hudson Alpha Institute for Biotechnology, Huntsville, AL.

As genetic information is insufficient to unambiguously implicate many disease-causal variants, annotations that enrich for causal variation are essential. Current annotations tend to exploit a single information type (e.g., conservation) and/or are restricted in scope (e.g., to missense changes). A broadly applicable metric that objectively weights and integrates diverse information is needed. Here, we describe Combined Annotation Dependent Depletion (CADD), a framework that integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations. We implement CADD as a support vector machine, trained to predict the disease causality of variants from 14.7 million simulated variants. We pre-compute CADD-based scores (C-scores) for all 8.6 billion possible single nucleotide variants of the reference genome and enable scoring of short insertions/deletions. C-scores strongly correlate with allelic diversity, pathogenicity of both coding and non-coding variants, and experimentally measured regulatory effects, and also highly rank causal variants within individual genome sequences. Finally, C-scores of complex trait-associated variants from genome-wide association studies (GWAS) are significantly higher than matched controls and correlate with study sample size, likely reflecting the increased accuracy of larger GWAS. Thus, the ability of CADD to quantitatively prioritize functional, deleterious, and disease causal variants across a wide range of functional categories, effect sizes and genetic architectures is unmatched by any current annotation and will be widely useful for the identification of causal variation in both research and clinical settings.

1473W


Recently, many methods have been developed for the association studies of rare variants. However, current rare variants studies have focused only on gene-level association with diseases. We investigate gene-set-level association in order to sum up the effect of rare variants on diseases from the same gene set. We first present a new unified quadratic test which is shown to be generally more powerful than or as powerful as other tests. Using this quadratic test, we then develop the gene-set-level association tests for rare variants. This gene-set test is capable of handling the different directions of effects. It can also perform association test for the data including both common and rare variants. Through simulation studies, the proposed test was shown to outperform other rare variants tests in terms of average power under the various scenarios. We applied the proposed gene-set test to discover the association between the lipid-related traits and 2.5 million variants from 10,000 exome sequencing data from five major ethnic groups (African American, East Asian, European, Hispanic and South Asian) in T2D-GENES consortium. Using 1452 canonical pathways from MSigDB (v3.1) database, we identified several known gene sets at the 5% significance level after Bonferroni adjustment, which are related to the lipid including asparagine(N)-linked glycosylation, lipoprotein metabolism, and HDL-mediated lipid transport pathways. Within these pathways, B4GALT2 and PLTP were identified as most significant genes.

1474T

Monitoring, analyzing, and exploring Ion Torrent™ NGS data with Torrent Suite™ Software. D. León. Ion Torrent by Life Technologies, South San Francisco, CA.

Ion Torrent™ Software is specifically designed to allow Ion PGM™ and Ion Proton™ Instrument users to monitor, analyze and explore Ion Torrent™ NGS data. This report illustrates how instrument users can plan a sequencing run that includes not only instrument settings, but it also enable users to preset configurations for data analysis, variant calling and exporting to other annotation software. The Torrent Suite™ Software is accessed through the Torrent Browser, the associated web-based interface for monitoring, planning, reviewing and managing sequence data in real time. As sequence data are being generated, users can monitor a sequencing run and the quality of their data. After a sequencing run is complete, the main data analysis step of the software includes base calling and mapping of the sequence data to a reference genome. Variant calling for homozygous/heterozygous SNPs and small/large indels is the next automated step in the data analysis process with interactive tables of identified variants may be visualized and evaluated using Broad Institute’s Integrative Genomics Viewer and two validation assay search sites (TaqMan® SNP Genotyping Assays and PCR/Sanger Sequencing Primers) on Lifetechnologies.com. The architecture of the software also enables users to create plugin applications that perform custom analyses and can connect to third-party providers. The plugins that are included in the Torrent Suite™ Software installation include: Alignment,CoverageAnalysis,ERCC_Analysis,FastQC,FilterDuplicates,IonReporterUploader,RunRecogniTON,SampleID,SFFCreator,FilterDuplicates,IonReporterUploader,RunRecogniTON,SampleID,SFFCreator,TorrentSuiteCloud, and variantCaller. In addition to providing data quality information and summary analysis reports, Torrent Suite™ Software provides customers the option to further annotate and filter their list of variants. Specifically, they can choose to run a pre-set workflow with Ion Reporter™ software after a sequencing run is complete. This step in the analysis can be pre-configured in the Torrent Browser when planning a sequencing run. This optional annotation step provides biological associations for the variants of interest and allows customers to extend the data analysis capabilities of Torrent Suite™ Software. In summary, this report demonstrates, how with a single web interface, Torrent Suite™ Software enables Ion Torrent™ instrument users the ability to start a sequencing run by setting up the sequencing parameters and the desired downstream analyses. (Research Use Only).
1475F
FERRET: a User Friendly Tool to Quickly Extract Data from the 1000 Genomes Project. S. Limou, G. Nelson, P. An, C. Winkler. Basic Science Program, SAIC, Frederick National Laboratory for Cancer Research, Frederick, MD.

By sequencing individual genomes from several reference populations, the 1000 Genomes Project (1KG) provides a valuable and near-comprehensive resource on human genetic variation for the scientific community. Raw and annotated data are regularly and rapidly released on an ftp server, and variants may be accessed directly through the 1KG browser. Even though the 1KG browser is well-designed, accessing data of interest can become a tedious process requiring both a good knowledge of the website architecture and many clicks. Data accessible from the 1KG browser are not updated as quickly as the data released into the ftp server. Further, the ‘VCF to PED’ tool available from the 1KG browser does not support indels nor provide variant frequencies. We developed Ferret, a user-friendly tool to quickly extract data from the latest release on the 1KG ftp server. Ferret is a Perl script parsing the 1KG vcf files to extract allelic frequency and genotype data for each variant (SNP and indel) located in the region of interest for the reference population of interest. Genotype data are recapitulated into map, ped and info files, which may be loaded in PLINK or HaploView for further exploration of linkage disequilibrium pattern, haplotypes, and eventually tagSNP design for customized genotyping arrays. The main advantages of Ferret are (1) the handling of indels, (2) the user-friendly interface, (3) the output format, and (4) calculation of allele frequencies. Ferret is thus a straightforward program, even for non-specialists who are not adept at the 1KG bioinformatics tools; this software permits easy manipulation and visualization of 1KG data with well-known pre-existing tools. Ferret is publicly available at: https://ccrnl.cancer.gov/confluence/display/BCGC/BCGC+Software. You may contact the first author (during and after the meeting) at sophie.limou@nih.gov.

1476W
dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. X. Liu, X. Jian, E. Boerwinkle.

Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX.

dbNSFP is a database developed for functional prediction and annotation of all potential non-synonymous single-nucleotide variants (nsSNVs) in the human genome. This database significantly facilitates the process of querying predictions and annotations from different databases/web-servers for large amounts of nsSNVs discovered in exome-sequencing studies. Here we report a recent major update of the database to version 2.0. We have rebuilt the SNV collection based on GENCODE 9 and currently the database includes 87,347,043 nsSNVs and 2,270,742 essential splice site SNVs (an 18% increase compared to dbNSFP v1.0). For each nsSNV dbNSFP v2.0 has added two prediction scores (MutationAssessor and FATHMM) and two conservation scores (GERP++ and SIFT). The original five prediction and conservation scores in v1.0 (SIFT, Polyphen2, LRT, MutationTaster and PhyloP) have been updated. Rich functional annotations for SNVs and genes have also been added into the new version, including allele frequencies observed in the 1000 Genomes Project phase 1 data and the NHLBI Exome Sequencing Project, various gene IDs for different databases, gene functional description, and disease-relevant sites, pathways, tissues/organs the gene expressed in, other genes the gene interacted with, estimated probability of haploinsufficiency/recessive-disease-causing, etc. A companion java program is provided for quick local query of SNVs/positions/genes with support of using a vcf file directly as input. dbNSFP v2.0 is freely available for download at http://sites.google.com/site/jpopgen/dbNSFP.

1477T

NCBI’s Conserved Domain Database (CDD) is a protein classification and annotation resource. Manually curated multiple sequence alignments (MSAs) are used to define families that are annotated with functional descriptions, links to the literature, and functional sites based on 3D protein structure or experimental evidence. Functional sites often have highly conserved sequence and/or structure and are closely linked to particular structural or functional roles. We investigated the presence of single nucleotide polymorphisms (SNPs) with human diseases at such functional sites and demonstrated how pre-computed sequence annotation available from CDD may assist in understanding phenotypes caused by sequence variation. We employ the medically important nuclear receptor (NR) family as an example for this analysis. Nuclear receptors are ligand-modulated transcription factors that act in the nucleus to regulate target gene transcription through the interaction of cofactor proteins. Abluent receptors result in a wide range of prevalent human diseases and disorders, such as cancer, diabetes, obesity. About 10% of the most prescribed drugs act through nuclear receptors, reflecting their medical importance. Nuclear receptors share a common structural organization with a central well conserved DNA binding domain (DBD), a variable N-terminal domain, a non-conserved hinge and a C-terminal ligand binding domain (LBD). We have classified LBD and DBD into approximately 50 unique subfamilies. The domain models representing these subfamilies were manually curated to add functional sites annotations, specifically ligand binding sites, DNA binding sites, co-activator/repressor interaction sites and residues involved in ligand/glycogenization. Genes corresponding to the human members of the NR family and non-synonymous SNPs collected for these genes were extracted from the NCBI’s Gene database. The SNPs were mapped to the domain footprints and functional sites annotated by CDD. We report on the correlations between genetic variations that overlap the annotated functional sites and human diseases.

1478F
H3M2: a novel algorithm for the detection of Runs of Homozygosity from second generation sequencing data. A. Magi1, L. Tattini1, P. Flicek1, M. Benelli2, M. Sen3, G.F. Gensini4, G. Romeo4, T. Pippucci5. 1) Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy; 2) U.O. Medical genetics Policlinico Sant’orosia matipghi, Bologna, Italy; 3) Diagnostic Genetic Unit, Careggi Hospital, Florence, Italy; 4) Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy.

Runs of Homozygosity (ROHS) are genomic stretches that appear in homozygous state, showing both alleles as identical in a diploid genome. ROHS are medically relevant. In rare autosomal recessive disorders (ARD), a homoygous highly penetrant mutation will probably reside in an unusually long homozygous haplotype originated from recent consanguinity. In complex disorders, recessive risk variants can be associated to shorter ROHS in inbred/outbred populations. To date, ROH discovery has been performed using microarray-based technologies. We introduce a novel computational approach based on heterogeneous hidden markov model, H3M2 (Homozygosity Heterogeneous Hidden Markov Model), for the identification of ROHS from HTS data. To develop this method, we used whole-exome sequencing (WES) data produced by the 1000 genomes project (1000 GP) consortium previously genotyped by the HapMap consortium. As a measure of homozygosity/heterozygosity for each polymorphic position i, we used the B-allele frequency (BAF) that is defined as the ratio between B-allele counts (NB) the number of reads that match with the allele with minor frequency at position i and the total number of reads mapped to that position (N, the depth of coverage). Then, based on the population allele frequency of all the polymorphic positions tested, we calculated for each detected homozygous region a so-called Regional Diploype Score (RDS) as a measure of statistical significance. To test the ability of our algorithm to identify ROHs of different size and/or number of SNPs, we applied it to WES data of six individuals sequenced by the 1000GP Consortium. H3M2 was able to identify more than 90% of the ROHS detected with HapMap SNP array data, outperforming state of the art methods. We also applied the algorithm to 100 1000GP individuals from different populations, showing that the total length of ROHS per individual increases with the distance of a population from East Africa, in agreement with previous observation. Finally, we tested the algorithm on 10 offspring to consanguineous families affected with ARD. In those subjects for whom the disease-mutation was known we found that the RDS of the surrounding region ranked among the highest subject’s ones (P D=0.05), demonstrating that the proposed algorithm is a valid tool also for disease-focus mapping. H3M2 is a highly performing algorithm aimed at the identification of differently sized ROHs from HTS data, in populations as well as in individuals.
1479W

Increasing volumes of next-generation sequencing data and the growing costs of experimental validation of detected changes call for the development of bioinformatics tools that can help prioritize genes for further analysis. We have developed a computational pipeline (PRIOR) that combines two published bioinformatics tools (snpeff and AnnTools) with additional methods obtained by our group. Variant annotation is performed with the snpeff and AnnTools and parsed with the AnnTools' parser. The effects with the highest impact are considered candidates for experimental validation, while all other effects are preserved for possible future studies. The pipeline accepts input data in the standard VCF format and annotates the variants to include only novel (not in the dbSNP), potentially disease-causing mutations in the coding regions, which include non-synonymous single nucleotide substitutions (SNP/SNV) and short insertions/deletions (INDEL). The tool makes predictions of the potential importance of each gene based on the overall mutation rate in a study cohort, and calculates the mutation rate for each gene adjusted, firstly, for its length and, secondly, for mutation rate in general population based on the rate of novel SNP/SNV and INDEL reported by NHLBI Exome Sequencing Project (ESP) in 6500 exomes. For studies that involve related individuals, further adjustment is made to add more weight to variants found in not related individuals compared with those found in the family members. Additional annotation for each gene includes function description, functional pathways, and known disease association as reported in the dbNSFP database. The application is freely available for public use; the package includes installation scripts and a set of helper tools.

1480T

Finding and accessing the ever-growing amount of genomic data and meta-data can be difficult for researchers and clinicians. Once a target data set is identified, extracting that data set from repositories and running analyses can tax or overwhelm available information technology resources such as compute power, storage and networks. By leveraging specialized software tools and an information technology infrastructure designed specifically for genomic data, commonly encountered challenges can be overcome, resulting in a shortcut to data analysis. As organizations struggle with increasing demands from researchers for access to high performance compute platforms, they often experience reduced throughput during peak times, coupled with hardware that is woefully underutilized during non-peak times. Network limitations and firewall issues, manifested as lengthy, error prone Internet downloads are common when migrating datasets from external systems to compute centers. At the compute centers, temporary storage can be inadequate for the ever-increasing amount of data that must be migrated in preparation for analysis, creating yet another obstacle to the utilization of the entire data. A shortcut is needed to overcome these ubiquitous problems in compute, storage and networks. One solution is setting up optimally architected data repositories co-located with high performance compute centers on efficiently networked. Annai Systems' web based computing platform plus GNOS web services is one such solution. Researchers can locate target datasets through customized searches in reQuest across multiple GNOS enabled repositories. The GTFuse access tool allows access to specific regions of data without downloading the whole genome sequences. For example, TCGA credentialed researchers can swiftly compile datasets and run analyses on Annai System's BioCompute Farm. This is a GNOS web based service that provides elastic compute and storage, and is co-located with the Cancer Genomics Hub, the largest cancer genome repository. Scientists using the Cancer Genome Atlas data via the reQuest portal and working in the BioCompute Farm with GTFuse are able to extract specific gene segments from hundreds of whole genome sequences in hours as opposed to the days or weeks it could otherwise take. We will discuss how scientists and clinicians are empowered by shortcuts to finding, accessing and analyzing genomic data through the use of the tools described above.

1481F
High Throughput Exome Coverage and Capture of Clinically Relevant Cardiac Genes. D. Manase1,2, L. D’Alessandro2, A.K. Manickaraj1,3, S. Mital1. 1) Heart Centre Biobank Registry. Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Cardiology, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada.

Background: High throughput whole-exome sequencing (WES) technologies have been used to identify variants in human disease. However, variant identification is limited by the overall coverage provided by WES. We sought to quantify the overall capture and read depth (RD) coverage of clinically relevant cardiac genes in WES data obtained using two of the most popular sequencing platforms. Methods: Data from 200 human exomes sequenced via Illumina HiSeq using Agilent SureSelect for capture were obtained in raw sequence format and aligned to the human genome version 19 using Burrows-Wheeler Aligner. Of these, 45 BAM files were analyzed using SAMTOOLS. Fifty clinically relevant cardiac genes were selected for analysis including 31 genes on the American College of Medical Genetics list for reporting of incidental findings and an additional 19 genes associated with congenital heart disease. Gene coordinates were obtained for all protein coding regions on Ensembl and filtered for unique chromosomal locations. Both the RD across all coding exons and the BED file coverage from the sequencing capture kit were compared with the reference genome to assess observed versus potential capture. Results: Preliminary results of 50 cardiac genes analyzed on 45 exomes sequenced at 50X coverage revealed that a RD coverage of 50X was seen only on 24% of all exonic regions. RD between 0 to 9 reads was the predominant coverage observed in these genes occurring at 41% of all exonic regions. Of note, none of the 50 genes had more than 70% of all isoforms within 50X of coverage. Only 1 gene had 100% of its isoforms covered at less than 10X reads. APOB and MYH11 were the only genes with at least half of all regions covered ≥ 5X, ACVR1, CFC1, GATA4, TSC2, NRAS, PTPN11, SMAD3, SOS1, TBX5, TGFB1, TGFB2, and TMEM43 had the least coverage in RD with 50% of the regions covered by only 0-9 reads. Conclusion: Low capture and RD depth may impact variant analysis and interpretation, including false negatives wherein potential heterozygous mutations may not present within the region that are sequenced while true calls may be ignored through low read depth filtering or even interpreted as artifact or misaligned read. A study bias towards well-covered genes may result in greater attribution of disease causation to mutations in well-represented genes with more confident calls. Improvements in WES technology are needed before widespread clinical use of exome data.

1482W

As the cost of sequencing continues to drop, the amount of bioinformatics data researchers need to manipulate and store increases exponentially, creating an ever-increasing demand for the computational methods required to manipulate data, perform statistical analyses, create graphics, and share results. For the majority of researchers this involves selecting the most relevant packages and either purchasing proprietary software or finding and compiling an appropriate open source equivalent. Maintenance of any such setup can become burdensome and the software support role can grow to dominate the time spent on actual research. We have developed a new and better method for quickly importing software applications to any Cloud computing provider such as Amazon AWS, HP Cloud, Rackspace, or Microsoft Azure, and have used this to create an Open Science Platform for use by researchers. This platform takes on the burden of compiling and supporting the common applications. Using the CIQr Open Science Platform researchers can share data and results with collaborators in an easy to use manner, without the need to compile software and concentrate on their investigations. The Open Science Platform can significantly reduce the analysis timeframe and cost to researchers. Pricing for the open source software is based upon the compute and storage resources consumed, reducing the barrier to use significantly and encouraging experimentation. We will discuss the technical challenges in the creation of this platform, issues around Cloud computing and bioinformatics, and how computation and storage requirements drive choice of Cloud. Use of multiple Cloud providers takes advantage of the diversity in Cloud representation will be demonstrated. The benchmarking capabilities of the Open Science Platform will be presented, and the rapid transition of applications between Clouds to allow for optimum performance or cost behavior.
1483T

Liftover of Short Reads from Next Generation Sequences Aligned to Non-Canonical References as a Step towards a Diploid Alignment.

T.C. Markello1,2, T.R. Gahl1, A.J. Brandt1, P.J. Pemberton1, R. Simeonov1, E.D. Flynn1, W.P. Bone1, A.E. Links3, D.R. Adams1,2, C.F. Boeroekool1, W.A. Gahl1,2,3, 1) Med Gen Branch, NIH/NHGRI, Bethesda, MD; 2) Undiagnosed Diseases Program, NIH, Bethesda, MD; 3) Office of the Clinical Director, NHGRI, NIH, Bethesda, MD.

Using a canonical reference genome for aligning short reads to a template is less computationally intensive than de novo assembly, but the alignment may be inaccurate if the reference is too dissimilar to the genome from which the reads were derived. Previous work has shown that it is possible to modify a canonical reference to more closely match the unknown sequence by applying corrections derived from hybridization experiments or imputed from population based haplotypes. The biggest unsolved challenge is to relate the alignment performed onto a modified reference back to the canonical reference. This is required in order to annotate the discovered variants for their significance to cause disease. A chain file contains all the information required for mathematical transformation of coordinates between references and is generated when modifying a genome sequence. If modifications are restricted to changing nucleotides from one base to another, then there is a one to one correspondence between reference positions. When modifications generate length discrepancies between the two sequences by including copy number variations, then positions in the modified reference must be lifted over back onto the canonical reference in chain blocks. After short reads are aligned to a modified reference their record includes a Compact Idiopathic Gapped Alignment Report (CIGAR) string that can span two or more chain file blocks and require modification to account for sequence differences. Using the Liftover Group of the Undiagnosed Diseases Program (UDP) has written and implemented an extension to existing bioinformatic suites that preforms the CIGAR string liftover function. We have extended this code to liftover the complex variant combinations arising from the correction of alignment references in a sequenced parent-parent-child trio. We present an analysis of the challenges associated with this lift-over process, including examples of local alignments that are improved by our current process. A successful liftover of short reads aligned to a modified reference allows the user to set up a workflow to go through the workflow is set up, it can be installed on the Workbench or on a Server. To another tool. In this way, the user can set up a workflow to go through a complete list of candidates for novel undiagnosed diseases using sequencing. This approach is particularly important when striving to generate a whole-genome sequencing study. We compare results using the RefSeq and Ensembl transcripts sets as the basis for variant annotation and find only 63% agreement in annotations for putative loss-of-function variants. The rate of matching annotations remains low for loss-of-function and nonsynonymous variants combined (81%) and all exonic variants (86%). Further, we compare the results from Annovar and VEP when using Ensembl transcripts and see matching annotations for only 85% of exonic variants, with particularly large discrepancies in the annotation of splicing variants. Using these comparisons, we characterize the types of apparent errors made by Annovar and VEP and discuss their impact on the analysis of DNA variants in whole-genome sequencing studies.

Our results show that variant annotation is not yet solved. Choice of transcript set has a large effect on the ultimate variant annotations obtained in a whole-genome sequencing study. Choice of annotation software has a smaller, but nevertheless important, effect. The annotation step in the analysis of a genome sequencing study must therefore be considered carefully, and a conscious choice made as to which transcript set and software are used for annotation.

1484F

Workflows for variation detection and filtering using NGS data from targeted resequencing, whole genome analysis, and RNA-Seq.

M. Matienko1, C. Boysen1, A. Joecker1, A. Joecker2, S. Monsted2, R. Forsberg2.
1) CLC bio, Cambridge, MA; 2) CLCbio, Aarhus, Denmark.

CLC Genomics Workbench provides a graphical and user-friendly framework for creating, distributing, installing, and running workflows. A workflow consists of a series of tools where the output of one tool becomes the input to another tool. In this way, the user can set up a workflow to go through multiple steps of analysis. For example, the read trimming output is used for mapping to reference, the mapping file is used for variant detection, and the resulting variant track is filtered against a standard variation track. Once the workflow is set up, it can be installed on the Workbench or on a Server. The workflow files are distributable and can be sent to other CLC bio users. Workflows can be run in batch mode like many of the Workbench tools, and allow for the analysis of multiple samples using the same pipeline. Any or all of the workflow parameters can be locked, so standard procedures can be used in production or regulated environments. Workflows created in CLC Genomics Workbench can also be installed on CLC Genomics Server. He re we show the example workflows created in Workbench for variation analysis on NGS data from human whole genome resequencing, targeted resequencing, and RNA-Seq data.

1485W

Choice of transcripts and software has a large effect on variant annotation.

D. McCarthy1,2, A. Kanapin2, M. Rivas2, K. Gaulton2, P. Humphrey2.
1) Department of Statistics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom.

Functional annotation of variants is a crucial step in the analysis of whole-genome sequencing data. In disease studies, functional annotation results can have a strong influence on the ultimate conclusion of the study. Incorrect or incomplete annotations can cause researchers both to overlook potentially disease-relevant DNA variants and to dilute truly interesting variants in a pool of false positives. Researchers are aware of these issues in general, but until now the extent of the dependency of final results on the choice of transcripts and software used for annotation has not been appreciated to the extent it should.

We quantify the extent of differences in annotation of 10 million variants from a whole-genome sequencing study. We compare results using the RefSeq and Ensembl transcripts sets as the basis for variant annotation with the annotation software Annovar, and find only 63% agreement in annotations for putative loss-of-function variants. The rate of matching annotations remains low for loss-of-function and nonsynonymous variants combined (81%) and all exonic variants (86%). Further, we compare the results from Annovar and VEP when using Ensembl transcripts and see matching annotations for only 85% of exonic variants, with particularly large differences in the annotation of splicing variants. Using these comparisons, we characterize the types of apparent errors made by Annovar and VEP and discuss their impact on the analysis of DNA variants in whole-genome sequencing studies.

Our results show that variant annotation is not yet solved. Choice of transcript set has a large effect on the ultimate variant annotations obtained in a whole-genome sequencing study. Choice of annotation software has a smaller, but nevertheless important, effect. The annotation step in the analysis of a genome sequencing study must therefore be considered carefully, and a conscious choice made as to which transcript set and software are used for annotation.

1486T

Joint GWAS Analysis Demonstrates Increasing Similarity among Disparate Diseases as Genomic Resolution Rises from SNP to Gene to Pathway.

M. McGeachie1,2, G. Clemmer1, J. Laskey-Su1,2, H.H. Chang1,2, B. Raby1,2, S. Weiss1,2, 1) Channing Div Network Med, Dept of Medicine, Brigham and Women’s Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Children’s Hospital Informatics Program, Children’s Hospital Boston, Boston, MA.

Network models of genetic diseases, polygenic modeling, clinical comorbidity and gene pleiotropy studies have led to a broader conception of the genetic causes of complex human diseases: one including increased relationships and interdependencies between disparate disorders. To assess the value of combining existing GWAS of separate cohorts and separate complex human genetic diseases for discovering additional disease-relevant biology, we compared six different GWAS from the WTCCC (bipolar disorder, cardiovascular disease, Crohn’s disease, rheumatoid arthritis, and type 1 and 2 diabetes). We developed a novel methodology called Joint GWAS Analysis that is based on the enrichment of top SNPs among two GWAS for different diseases. We used Joint GWAS Analysis to examine the similarity of diseases at the SNP level, gene level, and biological pathway level. We demonstrate that this new method identifies disease-relevant biology than could be obtained from single GWAS alone by comparing Joint GWAS Analysis and traditional single GWAS to the SNPs and genes reported in the NHGRI catalog of published GWAS for each disease. After running DAVID enriched pathway cluster analysis of the SNPs and genes reported in the NHGRI catalog genes for each disease, Joint GWAS Analysis identified significantly more pathway clusters than single GWAS, showing gains of 27% to 100% in bipolar disorder, cardiovascular disease, and type 1 and 2 diabetes. Such gains were not realized at the SNP and gene level. These results imply that complex human genetic disorders, such as those assessed by WTCCC GWAS, show broad similarity level in enriched biological pathways, even in cases of disorders not traditionally considered to have common genetic or etiological underpinnings. We also make concrete hypotheses regarding novel pathway associations for several of the complex diseases assessed by WTCCC, based on the results of Joint GWAS Analysis. Furthermore, the success of Joint GWAS Analysis indicates that meta-analysis of GWAS need not be limited to GWAS of the same phenotype.
Clinical - A variant analysis tool for next generation clinical sequence data. S. McGee, T. Kolar, M.O. Dorschner, J.D. Smith, D.A. Nickerson.

Next Generation Sequencing (NGS) has revolutionized molecular diagnostics. Test complexity and data volume have increased exponentially, creating significant challenges for clinical laboratories. Each test requires the comprehensive review of primary and variant data in a number of databases. To facilitate analysis, interpretation and reporting, we have developed an integrated tool to view sequence alignments, quality metrics and variant annotation. BAM and VCF files are easily uploaded, and variants filtered by user defined parameters and sorted by variant name, genomic position, and presence or absence in specific databases and functional annotations and ranking. This platform summarizes data for multiple samples simultaneously and links to existing databases to make quantitative decisions regarding validity and categorization of the variants. This is done by utilizing an extensive set of exome sequences to report a statistical measure of confidence for each newly discovered rare variant based on the quality of all calls previously seen at that site in more than 5,000 exomes. Data can be exported to one of the application’s summary templates to create comprehensive reports that include the reported variants and pertinent commentary that can be accessed. This tool provides an easy and efficient way to integrate a myriad of data into a single viewable format for clinical genetics.


We present a streaming algorithm which can find key statistics for the de Bruijn graph without constructing the de Bruijn graph. The algorithm requires only one pass over sequence data and uses a constant amount of memory, independent of the size of the dataset. The streaming algorithm can report the statistics to an arbitrary precision at the cost of increased memory usage. The statistics reported include the total number of k-mers and number of repeated k-mers. These statistics can be used to determine the per-basepair sequencing error rate without mapping to a reference. Additionally we can estimate average contig length for the de Bruijn graph, both when all k-mers are included and when unique k-mers are filtered.


Mutations in the mitochondrial genome are known to be associated with various pathological conditions and disorders caused by defects in components of the respiratory chain, including neurological disorders, maternally inherited diabetes & deafness (MIDD), among others. To improve clinical diagnosis and increasing the number of detected variants, a local Next Generation Sequencing (NGS) tool has been developed to analyze the whole mitochondrial genome (mtDNA). The definition and characterization of the analytical approach used to interpret the NGS data is a critical component of the test development process. Here we report on two critical analytical components related to the mitochondrial test development: cross-platform validation and large deletion detection. Using a dual-sequencing strategy, we sequenced ~35 mtDNA samples on both Illumina and Ion Torrent PGM instruments. MiSeq data was analyzed using the CLC Bio Server. For the Ion Torrent data, we evaluated two methods, using the Torrent Server plug-ins or using the CLC Bio Server, to identify optimal analytics. Our goal is to use one NGS platform as an orthogonal confirmation platform for variants detected on the other. Variants can occur selectively in any percentage of mitochondrial copies (heteroplasmy), which can contribute to high variability in disease severity. Given the high coverage of sequencing in our samples, we are able to observe as low as 5% heteroplasmy. To define our false negative rate, we also sequenced 12 Hapmap samples and compared the concordance of variant calls. Mitochondrial DNA deletion syndromes such as Kearns-Sayre Syndrome & Pearson’s Syndrome are caused by large deletions. To assess large deletion detection syndromes such as Kearns-Sayre Syndrome & Pearson’s Syn-

Whole Genome Sequencing (WGS) has the potential to transform diagnostic testing in very near future. However, it is computationally expensive to align millions of short reads to the whole genome. This could be prohibitive for the clinical setting where speed of analysis can impact patient outcome. Although the time consuming steps of WGS bioinformatics are highly computational, clinical applicability can be improved by prioritizing reports based on clinical decision-making. Current clinically relevant information is largely related to protein coding regions of the genome. Based on that observation, we have designed a sequence alignment and variant calling workflow focused on protein-coding exon-only regions before processing the data over whole genome. Our 2-step workflow significantly decreases turn-around time making it pertinent for clinical screening and diagnostics from WGS data. Our workflow first selects clinically relevant reads by aligning the entire sequencing data to the coding exons, limiting the initial reference to less than 2 percent of the whole genome. The resulting significantly reduced set of reads is then aligned to the whole genome to correct alignment artifact produced by the use of a reduced exont-only reference sequence. The data is re-aligned and re-calibrated followed by variant calling. The process thus uses a smaller reference sequence for all the sequence reads and then uses a smaller set of sequence reads to align to the whole genome. By using this two-step approach we are able to reduce the turn-around time for whole genome alignment and variant calling from ~80 hours to ~15 hours, a gain of more than 80 percent. The results are highly concordant to a standardly aligned WGS analysis apropos both the aligned reads as well as called variants. We also tested 3 samples from a well characterized HapMap trio. Variant calls were also validated against SNP array data. We present those results and metrics.

Locus Reference Genomic (LRG) record: reference resource for the reporting of clinically relevant sequence variants. J. Morales1, J.A.L. MacArthur2, R. Tully2, L. Gil2, A. Astashyn2, E. Bruford2, D. Dalgleish2, E. Birney1, P. Flicek1, D. Maglioti2, F. Cunningham1. 1) European Bioinformati-

Locus Reference Genomic (LRG) record is a manually curated resource designed specifically for the reporting of variants on clinically relevant geno-

Whole Genome Sequencing

Posters: Bioinformatics and Genomic Technology

Next-Generation Sequencing (NGS) has revolutionized molecular diag-

Whole Genome Sequencing

Next-Generation Sequencing (NGS) has revolutionized molecular diag-

Whole Genome Sequencing

Whole Genome Sequencing (WGS) has the potential to transform diagnostic testing in very near future. However, it is computationally expensive to align millions of short reads to the whole genome. This could be prohibitive for the clinical setting where speed of analysis can impact patient outcome. Although the time consuming steps of WGS bioinformatics are highly computational, clinical applicability can be improved by prioritizing reports based on clinical decision-making. Current clinically relevant information is largely related to protein coding regions of the genome. Based on that observation, we have designed a sequence alignment and variant calling workflow focused on protein-coding exon-only regions before processing the data over whole genome. Our 2-step workflow significantly decreases turn-around time making it pertinent for clinical screening and diagnostics from WGS data. Our workflow first selects clinically relevant reads by aligning the entire sequencing data to the coding exons, limiting the initial reference to less than 2 percent of the whole genome. The resulting significantly reduced set of reads is then aligned to the whole genome to correct alignment artifact produced by the use of a reduced exont-only reference sequence. The data is re-aligned and re-calibrated followed by variant calling. The process thus uses a smaller reference sequence for all the sequence reads and then uses a smaller set of sequence reads to align to the whole genome. By using this two-step approach we are able to reduce the turn-around time for whole genome alignment and variant calling from ~80 hours to ~15 hours, a gain of more than 80 percent. The results are highly concordant to a standardly aligned WGS analysis apropos both the aligned reads as well as called variants. We also tested 3 samples from a well characterized HapMap trio. Variant calls were also validated against SNP array data. We present those results and metrics.

Locus Reference Genomic (LRG) record: reference resource for the reporting of clinically relevant sequence variants. J. Morales1, J.A.L. MacArthur2, R. Tully2, L. Gil2, A. Astashyn2, E. Bruford2, D. Dalgleish2, E. Birney1, P. Flicek1, D. Maglioti2, F. Cunningham1. 1) European Bioinformati-

Locus Reference Genomic (LRG) record is a manually curated resource designed specifically for the reporting of variants on clinically relevant geno-

Whole Genome Sequencing (WGS) has the potential to transform diagnostic testing in very near future. However, it is computationally expensive to align millions of short reads to the whole genome. This could be prohibitive for the clinical setting where speed of analysis can impact patient outcome. Although the time consuming steps of WGS bioinformatics are highly computational, clinical applicability can be improved by prioritizing reports based on clinical decision-making. Current clinically relevant information is largely related to protein coding regions of the genome. Based on that observation, we have designed a sequence alignment and variant calling workflow focused on protein-coding exon-only regions before processing the data over whole genome. Our 2-step workflow significantly decreases turn-around time making it pertinent for clinical screening and diagnostics from WGS data. Our workflow first selects clinically relevant reads by aligning the entire sequencing data to the coding exons, limiting the initial reference to less than 2 percent of the whole genome. The resulting significantly reduced set of reads is then aligned to the whole genome to correct alignment artifact produced by the use of a reduced exont-only reference sequence. The data is re-aligned and re-calibrated followed by variant calling. The process thus uses a smaller reference sequence for all the sequence reads and then uses a smaller set of sequence reads to align to the whole genome. By using this two-step approach we are able to reduce the turn-around time for whole genome alignment and variant calling from ~80 hours to ~15 hours, a gain of more than 80 percent. The results are highly concordant to a standardly aligned WGS analysis apropos both the aligned reads as well as called variants. We also tested 3 samples from a well characterized HapMap trio. Variant calls were also validated against SNP array data. We present those results and metrics.

The emergence of high throughput sequencing of RNA has yielded a wealth of information regarding transcriptional regulation. Alignment of short read sequences to a common reference genome or transcriptome is a standard first step in the analysis of RNA sequencing data. We demonstrate that genetic variation away from the reference sequence can cause reads to be assigned to the wrong location. The source of most read alignment errors is in the duplicated structure of the transcribed genome and cannot be corrected by fine-tuning the alignment algorithm. We have developed a method, implemented as the software package Seqnature, to construct the imputed genomes of individuals (individualized genomes) of experimental model organisms including genetically unique outbred animals. Alignment to individualized diploid genomes increases read mapping accuracy and improves transcript abundance estimates. In an application to expression QTL mapping, this approach corrected erroneous linkages and unmasked hidden associations. Individualized genomes will be useful for other applications of high throughput sequencing technology that currently employ a reference sequence for alignment.

Bioinformatics and Data Management of High Throughput Sequencing Data from Prospective Cohort Study at Tohoku Medical Megabank Project. M. Nagasaki, S. Nariai, K. Kojima, Y. Yamaguchi, I. Sato, J. Yasuda, O. Tanabe, N. Fuse, K. Kenjo, R. Yamashita, Y. Yamagishi, I. Danjo, M. Matsumoto, K. Igarashi, K. Nakayama, F. Katsuki, S. Saito, I. Motaie, N. Ishida, M. Shiota, M. Yamamoto. Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan. Tohoku University Tohoku Medical Megabank Organization (ToMMo) was founded to establish an advanced medical system to foster the reconstruction from the Great East Japan Earthquake. The organization will develop a biobank that combines medical and genome information during the process of rebuilding the community medical system and supporting health and welfare in the Tohoku area. The information from the brand-new biobank will create a new medical system, and, based on the findings of its analysis, the organization aims to attract more medical practitioners from all over the country to the area, promote industry-academic partnerships, create employment in related fields, and restore the medical system in Tohoku. A blueprint for Tohoku University Tohoku Medical Megabank Organization is a ten-year project including three main activities: a biobank combining medical and genome information; an online platform for the coordination of community medical information; and training program designed for a variety of specialized professionals and experts such as researchers of bioinformatics and science communicators. The biobank to be developed will be utilized to analyze the local heredity information so that it can establish an advanced medical system based on genome information with cutting-edge information and communication technology. The first goal of ToMMo is to understand the genetical detailed background of population in this area including population specific rare variants. Thus, we will apply deep coverage whole genome sequencing of thousands people who applied to this prospective genome cohort project within years. This poster presents the poster main presenter's responsible part; the data management and the analysis of massive amount of high throughput sequencing data on this project and the research position availability of this very exciting project as a graduate student or a research staff.
1496F
CIDRVar: A Next-Generation Sequencing Database Linking Samples, Variants, and Annotations. J.D. Newcomer, S.M.L. Griffith, E. Pugh, D.R. Leary, J.J. Goldstein, L. Watkins Jr., K.F. Doehny, JHU/Center For Inherited Disease Research 333 Cassell Drive Triad Technology Building, Suite 2000 Baltimore, MD 21224. The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and statistical genetics consultation to investigators working to discover genes that contribute to disease. In April 2011, CIDR’s software development team implemented an ANNOVAR aggregation report generator as part of our software toolbox, CIDRSeqSuite. This report aggregation tool automates the process of running ANNOVAR over a directory of VCF files, annotating these samples against an arbitrary number of flat-file databases, and aggregating the output files into summary reports. Along with several custom databases, we use the gene-, region-, and filter-based annotations provided with ANNOVAR. This tool has been in production for over two years; in that time, we have accumulated more than 50 databases against which users may annotate based on their preferences. The accumulation of many very large annotation databases has caused scalability issues. To solve these problems, we have implemented our own solution: CIDRVar. A software tool backed by a relational database, CIDRVar serves two interdependent purposes: to track the variant genotypes found for all of our sequencing samples and to store data against which these variants can be annotated. Internally tracking the variants found for all of our sequencing samples allows us to answer quickly and easily important questions such as: Which samples had a variant at a given genomic position? Which variants were found at a given genomic position? What are the annotations for a particular variant or genomic location? How many samples had each variant? Which samples share a given genotype? Initially, report generation required server-grade hardware. Now that we are storing in a relational database all of the annotation data that were previously stored as flat-files, we can generate a variant annotation report with a much smaller resource footprint.

1497W
Improving the accuracy of novel sequence search in de novo human genomes with NSIT. B. Papadopoulous, A. Javed, M. J. Zaki, M. Ruchirawat. 1) Chulabhorn Research Institute, 54 Kampaeng Phet Rd. Laksi, Bangkok, Thailand 10210; 2) Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01, Singapore 138672; 3) Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York, USA 12180-3590. The idea of building the human pan-genome was first introduced in Li et al (2010). The authors estimated that each individual’s whole genome harbors about 5 Mb of novel DNA sequences that are neither known repeats nor present in the human reference sequence. These were shown to be potentially functionally important and consistent with known human migration paths. When combined, the complete human pan-genome is gauged to contain as much as 19–40 Mb of novel sequences. Subsequent independent research findings also identified 2-3 Mb of novel sequences per individual via different computational techniques. In this work, we followed up that about 1 Mb of novel sequences per individual is not as high as previously anticipated, i.e., only around 1.7–2.0 Mb per person, 2) the novel sequences among different individuals largely overlap, and 3) DNA sequence contaminations, such as artifacts due to PCR in a de novo assembly, result in low novel genome coverage detected as novel sequences. We used our software NSIT (Novel Sequence Identification Tool) to align individually de novo human genome assemblies (NA18507, YH, and NA12878) to the GRCh37 reference assembly. Repeats were masked with RepeatMasker and BLASTn was used for the final refined search steps. We found 1.9 Mb, 2.0 Mb, and 1.7 Mb of novel sequences in the aforementioned de novo assemblies, respectively. A further investigation revealed that these sequences are 60-80% overlapped among themselves as well as with the HgRef and CHM1 assemblies. Lastly, we might expect that as high as 129 kb of our NA18507 novel sequence candidates did not match with any existing human genome assemblies, or any other closely related species, but aligned with extremely high confidence and near perfect sequence identity to the Epstein-Barr virus (EBV) genome instead. More importantly, these 129 kb were also included in the novel sequences of NA18507 reported in Li et al. We speculate that these DNA sequence contaminations are present because EBV was used as the transformant when generating the cell lines. Our results therefore suggest that the size of the human pan-genome may not have as many sequence contaminants as expected. NSIT is a highly efficient and accurate software for the task of detecting novel sequences in a large de novo genome assembly. For the above experiments with human genomes, NSIT required <2GB of memory and ~2 hours on a commodity desktop. To the best of our knowledge, it is the only software designed specifically for this task.

1498T
ScatterShot: a Java program for creating cluster plots from Affymetrix and Illumina genotype data. N.W. Rayner, R. Robertson, M.I. McCarthy. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Oxford, United Kingdom; 3) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 4) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom. With the increased interest in rarer variants (<1% minor allele frequency) the number of these polymorphisms appearing on the newer genotyping arrays is growing. This has grown dramatically. This growth, and the uncertainty around the efficacy of calling of these variants, has greatly expanded the number of genotype cluster plots that need to be examined. To address this issue we have developed a cluster plotting program, ScatterShot, that is simple to run and is more flexible than existing programs in the data input formats required and display options available. ScatterShot is a Java program run from the command line, supporting multiple OS platforms (Windows, Unix, Mac OSX), and makes use of Java’s asynchronous IO libraries to facilitate parallel processing and increase throughput: the program will scale to make use of any available processors. A wide variety of input formats are supported such as the binary .chp files from Affymetrix’s genotyping console as well as the Final Report from Illumina’s GenomStudio. Also supported is a generic mode, taking genotype calls from a plink format ped or binary ped file coupled with a separate XY coordinate data table. The format of the XY data table is flexible allowing one row per SNP or one row per sample. An advantage of the program over others is that when using these multiple files, the ordering of the SNPs in the input files is not important. Outputs include the binary ped and chp files can be read from gzipped versions. Plots are output using Scaleable Vector Graphics (SVG) which can be viewed directly using modern HTML5 web browsers, or statically rendered into images files (JPEG, TIFF) or PDF files for printing. As an XML-based format, SVG also offers the opportunity to embed other annotations such as gender, cohort and batch information. Once embedded the SVG plot can be controlled dynamically using interactive JavaScript controls allowing for dynamic filtering of the displayed genotype calls. This can be, for example, removing samples with no genotype call assigned, or all samples from one cohort out of many. The plots use a consistent colouring tied to the genotype call and include basic SNP QC metrics such as HWE, call rate and heterozygosity. To date ScatterShot has been used for assessing the quality of novel clusters from the new Illumina Exome and Affymetrix Axiom BioBank chips both of which contain a high number of low frequency SNPs.

1499F
Functional annotation of non-coding variants. G.R.S. Ritchie, P. Flicek, E. Zeggini, UK10K Consortium. 1) European Bioinformatics Institute, European Molecular Biology Laboratory, Hinxton, Cambridge, Cambridgeshire, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, Cambridgeshire, United Kingdom; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 4) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom. Identifying functionally relevant variants against the background of ubiquitous genetic variation is a major challenge in human genetics. For variants that fall in genomic regions our understanding of splicing and the genetic consequences is fairly well understood but there are several groups of non-coding variants that are likely to affect protein function. There are, however, currently few methods to interpret variants that fall outside of coding regions and yet these are increasingly being identified as causally relevant in human disease. Efforts such as ENCODE and the Roadmap Epigenomics Project are producing a wide range of annotation in non-coding regions but it is not yet clear how to integrate these data into variation studies. To establish if any of these annotation methods can be used to interpret variants, we use the wealth of rare variants discovered in the whole genome sequencing arm of the UK10K project to quantify constraint in annotated regions across the genome, including non-coding genic elements, transcription factor binding sites and DNase1 hypersensitive sites. We identify a number of annotations, such as binding sites for the transcription factors BRF1 and BD1 that appear to be under comparable constraint to coding sequence (mean derived allele frequency (DAF) 0.0414 (95% CI: 0.0359-0.0489) and 0.0457 (95% CI: 0.0412-0.0502) respectively, compared to 0.0379 (95% CI: 0.0374-0.0383) for coding regions and a genome-wide average of 0.0676 (95% CI: 0.0676-0.0677), Wilcoxon rank-sum p < 1e-8 in both cases comparing the DAF of variants falling in the binding site with variants not falling in these annotations). We use the results from this analysis to develop a classification system of annotations that can be used to categorise non-coding variants in the genome, with a large positive predictive value for functionally important variants. We validate our approach using a large independent data set from a case–control association study looking at sequence changes in loci associated with type 2 diabetes.

Given the recent development of next-generation sequencing technologies on personal genomes, functional interpretation of genetic variants has become one of the major obstacles faced by biomedical researchers, clinicians and individual consumers wishing to analyze whole genome/exome sequencing data. Many research laboratories are flooded with high-throughput sequencing data sets, but the limited ability to process these data sets significantly delays the progress to infer biological insights. More and more consumers and patients are now sequencing their own genomes, but they often lack the ability to understand the health-related information buried under the variant calls provided by sequencing companies or command-line software tools. Although several software tools (such as ANNOVAR) have been developed, they suffer from a number of limitations, including the lack of a user-friendly interface, the lack of filtering and annotation workflows to analyze variant data. In particular, which is a graphical user interface for executing complex workflows on a command-line software tools. Although several software tools (such as ANNOVAR, which are executed on a high-performance cluster. The key

1502F Fast and high-resolution evaluation tool of Illumina high throughput data considering spatial organization of the sequencing clusters. Y. Sato, K. Kojima, N. Nariai, Y. Yamaguchi-Kabata, M. Nagasaki. Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan.

We developed a tool for fast and high-resolution evaluation of Illumina sequencing data by considering spatial organization of the sequencing clusters on a flowcell. The tool was designed as a part of quality control pipeline of the Tohoku Medical Megabank Project, and analyzes the sequence data generated by the Illumina HiSeq2500 High Output mode (HCS 2.0.05) and Rapid Run mode (HCS 2.0.0), and MiSeq v2 flowcells. This tool takes fastq, Casava fastq, and SAM/BAM files as input, and outputs heatmap plots of base quality per each cycle, read density, and mapping quality. The heatmap arrangement corresponds to the shape of the flowcell used (automatically detected). Each tile within the flowcell is divided up to 100×100 resolution and shown as respective cells of the heatmap plot. The heatmap visualizes technical errors of the sequencing, e.g., air bubbles and cracks in the flowcell, providing useful information for the improvement of sequencing procedure. In addition, this tool can convert to N base from nucleotide bases obtained from low-quality spots within the tile/flowcell. Such data processing based on the errors of flowcell or sequencing procedure would highly improve the downstream analyses, such as exploration of TF binding sites or low-frequency somatic mutations that requires high-quality read and mapping information.
deleterious variants. Although our performance estimates are likely overestimated, our methods (e.g. SIFT, Polyphen, MutationTaster) are still likely to be more efficient than classical deleteriousness prediction methods (e.g. ROC-AUC) by at least 30% compared to classical deleteriousness prediction methods. This is likely due to eXtasy's ability to discriminate between phenotype-specific and phenotype-unrelated deleterious variants. Although our performance estimates are likely overestimated, none of the prior information bias in a retrospective benchmark, we show that even controlling for these biases we obtain a substantial performance increase. We believe that the presented approach will greatly facilitate the analysis of exome sequencing data in human disease by efficiently prioritizing deleterious variants.

PHEVOR overcomes this limitation by propagating phenotype information across the Gene Ontology, extending Phenomizer's candidate list to more than 18,000 human genes, and (3) VAAST a probabilistic disease-gene finder integrating nSNVs in the light of the phenotype in question. eXtasy is publicly available at http://homes.esat.kuleuven.be/~bioiuser/eXtasy/.

PHEVOR: Integration of VAAST with Phenomizer and the Gene Ontology for accurate disease-gene identification using only a single affected exome. M. Singleton, L. Jorde, M. Yandell. Human Genetics, University of Utah, Salt Lake City, UT.

Accurate identification of disease-causing genes using only a single exome is a major analytic challenge. For recessive rare diseases, when only a single affected individual's exome is available, VAAST and AnnoVar typically rank the disease-causing gene among the top 10 candidates only 15% and 2% of the time, respectively. To improve disease-gene identification, we have developed PHEVOR, an algorithm that integrates three popular tools: (1) Phenomizer, which provides powerful means to generate candidate disease-gene lists based upon phenotype data, (2) the Gene Ontology (GO), which provides function, location and biological process descriptions for more than 18,000 human genes, and (3) VAAST a probabilistic disease-gene finder that uses variant-frequency information together with amino acid substitution (AAS) frequencies to identify disease-causing variants in personal genome sequences. We benchmarked PHEVOR using 50 known recessive disease-causing variants spiked into otherwise healthy exomes. On this dataset, VAAST identifies the disease-gene as the top candidate only 2% of the time and places it in the top 10 candidates 15% of the time. Phenomizer does about as well, ranking the disease-gene first 3% of the time and in the top 10 candidates 14% of the time—but with an important caveat: Phenomizer's candidate list is restricted to 18,000 genes in 30 complete genomes. If the disease-causing alleles do not reside in one of these genes, they will be missed as we have not prioritized the phenotype and genomics data. PHEVOR overcomes this limitation by propagating phenotype information across the Gene Ontology, extending Phenomizer's candidate list to more than 18,000 human genes. PHEVOR also provides probabilistic approaches to incorporate personal genome data using VAAST's proven statistical methodology. The improvement is dramatic. Using the same benchmark dataset described above, (note that 54/100 genes are not included in Phenomizer's gene list) and only a single affected exome, PHEVOR is able to identify 65% of disease-causing genes as the top candidate genome-wide, and 79% of the disease-gene is within the top 10 candidates. PHEVOR thus provides an effective new means for single exome-based diagnosis.

1504T


Massive parallel sequencing greatly facilitates the discovery of novel disease genes causing Mendelian and oligogenic disorders. However, many mutations are present in any individual genome, and identifying which ones are disease causing remains a largely open problem. We introduce a novel computational approach, called eXtasy, to prioritize nonsynonymous single nucleotide variants (nSNVs) by integrating variant impact prediction, haplinsufficiency prediction and phenotype-specific gene prioritization that allows significantly improved prediction of disease-causing variants in exome sequencing data. To train our method we use the Human Gene Mutation Database (HGMD) as our source of disease-causing variants and 3 control sets ranging from common polymorphisms to rare variation in healthy individuals. By integrating phenotype-specific gene prioritization information we are able to greatly increase the area under the receiver-operator curve (ROC-AUC) by at least 30% compared to classical deleteriousness prediction methods (e.g. SIFT, Polyphen, MutationTaster). This is likely due to eXtasy's ability to discriminate between phenotype-specific and phenotype-unrelated deleterious variants. Although our performance estimates are likely overestimated, due to prior information bias in a retrospective benchmark, we show that even controlling for these biases we obtain a substantial performance increase. We believe that the presented approach will greatly facilitate the analysis of exome sequencing data in human disease by efficiently prioritizing deleterious nSNVs in the light of the phenotype in question. eXtasy is publicly available at http://homes.esat.kuleuven.be/~bioiuser/eXtasy/.

1506W

Multiagent-based SNP Annotation for Large-Scale Genetic Diversity Analyses: An Application to the Brazilian EPiGEN Initiative. G.B. Soares-Souza, E. Tarazona-Bastos, A. Pereira, M.L. Barreto, B.L. Horta, M.F. Lima-Costa, A. Horimoto, N. Esteban, F. Khedly, W.C.S. Magalhaes, M.R. Rodrigues, The Brazilian EPiGEN Consortium. 1) General Biology Department, Federal University of Minas Gerais, Belo Horizonte, Brazil; 2) Laboratory of Genetics and Molecular Cardiology, Heart Institute, Medical School of University of São Paulo, São Paulo, Brazil; 3) Instituto de Saúde Coletiva, Federal University of Bahia, Salvador, Brazil; 4) Universidade Federal de Pelotas, Pelotas, Brazil; 5) Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil.

Despite the great amount of genetic diversity data available, the study of variants with biomedical interest requires their association with other types of biological data, such as pathways, pharmacogenetics, GWA, and so on. Since these heterogeneous data are fragmented in different databases, it is necessary to develop tools to integrate different sources of biological data in order to move forward with genetic studies of diseases or populations. Although there are annotation tools that solve this problem, the management of large scale data is still a challenge in terms of execution time. Computer technologies that incorporate parallel execution of the code are a promising solution for processing large scale data, since they are able to reduce execution time by dividing the execution load into independent processing units. One such technology is the agent-based technology. Here, agents are autonomous programs that perform a particular task and can communicate with other agents in a Multiagent system to delegate tasks, share results or get complementary information. Agents in a Multiagent system execute concurrently, as independent threads. Given the heterogeneity of the data and the distribution over distinct sources, we have chosen this technology to development a SNP annotation system for large scale data. In our system, Database Agents encapsulate databases of interest, such as dbSNP, pharmGBK, OMIM, UCSC and GO. The annotation process works as follows: (i) a Coordinator Agent provides a list of SNPs to be annotated and passes them on to each Database Agent; (ii) the latter Agents then retrieve, concurrently, relevant information from their encapsulated database to annotate the list of SNPs and return it to the Coordinator; (iii) after the annotated SNPs are retrieved, the Coordinator Agent merges the annotations and passes them on to the Interface Agent to generate an annotation report. Our Multiagent annotation system is being used to thoroughly annotate data from the Brazilian EPiGEN initiative, which involves the analysis of 1,238,813 SNPs in 6,600 individuals and 4.3 M SNPs in 270 individuals and 30 complete genomes. So far, tests showed that the distributed architecture of the Multiagent system significantly reduces the annotation execution time. As a follow up study, we will investigate the population genetics of different SNP classes to understand the implications of its allelic spectrum on GWAS studies, and new GxG and GxE approaches.
1507T

Comprehensive network and pathway analysis of RNA sequencing of triple negative breast cancers. J.P. Solzik1, R. Atalé 2, B.A. Hancock1, J.N. Billault2, M. Radovich1, 1) Surgery, Indiana University School of Medicine, Indianapolis, IN; 2) Ingenuity Systems, Redwood City, CA.

Introduction: Triple-negative breast cancers (TNBCs) account for 15% of all breast cancers cases and are defined by an absence of actionable therapeutic targets (ER, PR, HER2). Using RNA-seq data, we compared TNBCs to microdissected normal breast epithelium from healthy volunteers and to normal tissue adjacent to tumor followed by comprehensive network and pathway analysis. Methods: RNA-seq data from 34 TNBCs (from Indiana University Medicine Microdissection Tissue Bank), and 10 adjacent normal tissues (TCGA), were merged and imported into Partek Genomics Suite. The merged transcript RPMKMs were transformed, batch effect corrected, and analyzed for differential expression. Statistically significant genes were imported into Ingenuity Pathway Analysis (IPA) Spring 2013. Results: IPA analysis of differentially expressed genes of TNBCs compared to microdissected normal breast tissues identified key pathways in DNA damage, cell cycle, and immune signaling. We then employed Upstream Regulator and Causal Network Analysis which predict the activation or inhibition state of regulators based on downstream differentially expressed genes and known directionality of expression. From this analysis, two statistically significant upstream regulators with known roles in metastases were identified (interferon regulatory factor), and FOXM1 (cell-cycle transcription factor). In particular, FOXM1 is predicted to drive the activation of a multitude of genes involved in cancer cell proliferation. In addition, we identified two causal networks centered on inhibited SPDEF (Ets transcription factor), and on activated HOXB4 (developmental transcription factor). Such analysis has not previously been described in TNBC and may play a role in the stem cell phenotype of TNBC. In contrast, when TNBCs were compared to adjacent normal tissue, some of the same pathways and regulators were identified, but other pathways involved in fibrosis, atherosclerosis, metabolism, and edema were also observed, likely secondary to the stromal nature of adjacent normal tissue. Conclusion: By using RNA-seq data from TNBCs and microdissected normal epithelium coupled with comprehensive network and pathway analysis we demonstrate the utility of these methods to uncover novel biological processes involved in TNBC biology. In particular, we observe that IPA’s Causal Network Analysis has unique capabilities to identify master regulators useful for drug target discovery.

1508F

GrabBlur - a framework to facilitate the secure exchange of whole-exome and -genome sequencing data sets. B. Stade1, D. Seelow1, A. Franke1, 1) Institute of Clinical Molecular Biology, University Hospital Schleswig-Holstein, Kiel, Germany; 2) NeuroCure Clinical Research Centre, Charité - Universitätsmedizin, Berlin, Germany.

With ‘GrabBlur’ we developed a tool to collect, aggregate and share SNV (single nucleotide variant) data of hundreds of samples with a special trait/phenotype in public database, while keeping each individual sample unidentifiable. First, ‘GrabBlur’ helps a submitter to combine SNV data from samples with phenotype and sequencing information (VCF files) using a local web interface or command line operations. Examples are HPO terms (Human Phenotype Ontology), sample traits, information about the applied sequencing technology and the submitting institute / laboratory. Most of the information is optional - the submitter can decide for himself what he wants to and can share. In the second step ‘GrabBlur’ merges and aggregates the data. For example, if it turns out that a SNP is rare - every submitter sets the proper threshold for his own data - all sample information is replaced with contact information of the submitter. In further steps the data will be highly aggregated so it can be shared while adhering to data privacy. It will not be possible to recover the exome/ genome of an individual and family based quality metrics across all algorithms. The CGES approach is shown to outperform its constituents parts in many key quality metrics without a significant loss in the number of variant sites called. In particular we are able to achieve a threefold reduction in Mendelian inconsistencies between the best performing variant caller and our consensus approach (CGES = 240.14/trio and Atlas2.0 = 699.59/trio). For callers with comparable VAF and CAL scores, our CGES set of variants has an average QUAL score 11% (GATK) and 70% (Freebayes) higher than the unfiltered output set of each respective variant caller. Additionally, the consensus set outperforms all individual callers in the study with regard to expected exome-wide transition-transversion ratio (CGES = 3.07 and Atlas2.0 = 2.98). For the purpose of accessible, efficient and reproducible analysis we provide implementation of CGES as a stand alone command line tool, as well as a set of parallel Galaxy tools and workflows for accessible and efficient use by the research community (see Implementing a High Performance, Reusable Consensus Calling Pipeline for Next Generation Sequencing using Globus Genomics, Madduri el al., ASHG 2013).
**1511F**

**DRAW+SneakPeek: Analysis Workflow and Quality Metric Management for DNA-Seq Experiments.** O. Valladares 1,2, S. Ferreira 1, C.-F. Lin 1, Y. Yang 1, E. Chilldresw 1, I. Vellek 1, T. Geller 1, Y.-C. Hwang 2, E.A. Tsao 3, 4, D.D. Schellenberg 1, L.-S. Wang 1,2, 1) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA; 3) Department of Physics, University of Washington, Seattle, WA; 4) Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA; 5) Department of Pathology and Laboratory Medicine, The Children’s Hospital of Philadelphia, Philadelphia, PA.

We report our new DRAW+SneakPeek software for DNA-seq analysis. DRAW (DNA Resequencing Analysis Workflow) is based on commonly used open source programs for analyzing sequencing experiment data. DRAW implements the Best Practice Variant Detection from Genomic Analysis Toolkit (GATK) and automates the workflow of processing raw sequence reads including quality control, read alignment, and variant calling on High-Performance Computing (HPC) facilities such as Amazon Elastic Compute Cloud (EC2). SneakPeek provides an effective interface for reviewing dozens of quality metrics reported by DRAW, so users can assess the quality of data and diagnose problems in their sequencing procedures. We evaluated DRAW on Amazon EC2 using a single flowcell WES dataset of 350.2G nucleotides from 34 multiplexed samples using Nimblegen SeqCap EZ Human Exome Library, 100bp pair-end reads on a HiSeq2000 sequencer. DRAW processed the whole dataset in 1,943.2 core-hours, or 17.4 hours/core on 14 Quadruple Extra Large Instances with 112 cores, generating 1.1TB of data. The total cost was $528 including storage, computing, and data download. Both DRAW and SneakPeek are freely available under the MIT license. All programs and detailed documentation can be downloaded directly on Amazon Cloud with minimal installation. Instructions and codes can be obtained from the NIA Genetics of Alzheimer’s Disease Data Storage Site (NIAGADS) at [https://www.niagads.org/content/drawsnakepeek](https://www.niagads.org/content/drawsnakepeek).

**1512W**

**Performance of two imputation methods on large scale data: experiences in the eMERGE network.** S. Verma 1, G. Armstrong 1, M. Ritchie 2, D. Crawford 2, Y. Bradford 3, M. Andrade 1, I. Kullo 4, G. Trong 5, H. Kuiveniemi 6, L. Armstrong 7, G. Hayes 8, B. Keating 2, D. Crosslin 9, G. Jarvik 1, B. Namjou 1, E. Bookman 1, R. Liu 10, eMERGE Network. 1) Center for Systems Genomics, The Pennsylvania State University, University Park, PA; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Mayo Clinic, Rochester, MN; 4) Geisinger Health System, Danville, PA; 5) Northwestern University, Chicago, IL; 6) Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, PA; 7) University of Washington, Seattle WA; 8) Cincinnati Children’s Hospital Medical Center, Cincinnati, OH; 9) Division of Genomic Medicine, National Human Genome Research Institute.

The eMERGE network, an NHGRI funded initiative comprises nine sites each with DNA biobanks linked to electronic health records (EHRs). Approximately 39,206 unique DNA samples have been genotyped using either Affymetrix or Illumina genome-wide SNP arrays. Led by the Coordinating Center and the eMERGE genomics workgroup, we have developed an imputation pipeline for merging genomic data across the different SNP arrays using MAGIC (Maximum Allowable Gain in Imputation) to maximize sampling power to detect associations. We performed imputation using the 1000 Genomes Cosmopolitan reference panel - which includes 1092 individuals and over 36 million SNPs. We compared accuracy of imputation results from two software packages - Beagle and ImpuTe2 (phasing performed with ShapelIt2). For the comparison we used the following metrics: accuracy of imputation, allelic R2 (estimated correlation between the imputed and true genotypes for all imputed SNPs), and relationship between allelic R2 and minor allele frequency across all imputed SNPs. Since imputation is computationally intensive, we compared computation time and computing resources required by the two software packages. We have also outlined the major challenges and lessons learned due to the complexity of using these two approaches at different platforms. Finally, we will present lessons learned as we conducted our own imputation control process, as many of the pipelines are unique and cleaning of imputed data is essential for further analysis. This imputed dataset will serve as a valuable resource for variant discovery, leveraging the wide range of medically relevant phenotypes that can be mined from the EHR.

**1513T**

**FILTUS - a versatile and user-friendly program for filtration and statistical evaluation of variants in high-throughput sequencing projects.** M. Vigeland. Medical Genetics, Oslo University Hospital, Oslo, Norway.

We present FILTUS, a program for working with variant files resulting from sequencing projects, e.g. exome sequencing. Various options for filtering and summarizing statistics are available, as well as exploratory tools like pairwise sharing among the samples. Several statistical methods for evaluation of variants are implemented, aiding the identification of causal variants in Mendelian disease projects. Both case-control and family based sequencing designs are supported, as well as dominant/simple recessive/compound heterozygous regions. Homozygous regions can be identified (via PLINK), and the program offers conversion of variant files to MERLIN format, facilitating linkage analysis. FILTUS accepts variant files in any format, as long as the user identifies essential columns like chromosome, position, gene name, and type. Files can be loaded and saved in a user defined format and given special treatment. The implementation allows several hundred complete exomes to be analyzed simultaneously on a standard laptop. FILTUS has an easy-to-use GUI accompanied with a user manual, and does not require particular bioinformatic skills. It is written in Python and runs on Windows, Mac and Linux.

**1514F**

**DeMix: Deconvolution for Mixed Cancer Transcriptsomes Using Measured Data.** W. Wang 1, J. Ahrn 2, Y. Yuan 1, 1) Bioinformatics & Comp Biology, Univ Texas MD Anderson Cancer Ctr, Houston, TX; 2) Biostatistics, Univ Texas MD Anderson Cancer Ctr, Houston, TX.

Tissue samples of both tumor and stromal cells cause underestimation of gene expression signatures associated with cancer prognosis or response to treatment. In silico dissociation of mixed cell samples is essential for analyzing expression data generated in cancer studies. Currently, a systematic approach is lacking to address three challenges in computational deconvolution: 1) violation of linear addition of expression levels from multiple tissues when log transformed microarray data are used; 2) estimation of both tissue proportion and tumor-specific expression, when neither is known; and 3) estimation of expression profiles for individual patients. We have developed a statistical method for deconvolving mixed cancer transcriptsomes, DeMix, to address the above issues in array-based expression data. We demonstrate the performance of our model in synthetic and publicly available real datasets. Our method can be applied to ongoing biomarker-based clinical studies, as well as to the vast expression datasets previously generated from mixed tumor samples.

**1515W**

**üRRBS-Predictor, a webtool to guide enhanced-representation bisulphite (RRBS) analysis of genomic DNA methylation patterns.** T.R. Ward 1,2, X. Zhu 1,2, A.E. Urban 1,2, 1) Genetics Dept, Stanford, Stanford, CA; 2) Dept of Psychiatry, Stanford, Stanford, CA.

üRRBS-Predictor is a bioinformatics tool that predicts which parts of a given genome sequence will be included in an RRBS experiment if a given size-fraction of DNA fragments is chosen during the RRBS sample preparation. Genome-sequencing after bisulfite conversion is a powerful approach to understanding genomic DNA methylation. But whole genome bisulphite sequencing is costly and inefficient at today’s sequencing and computing speeds, due to GC poor regions in many genomes that will be included in the log transformed micromarray data. Deconvolution at all in terms of insight regarding the question at hand. RRBS uses restriction cutting at recognition sites containing CpG dinucleotides (typically with the restriction endonuclease MspI) followed by size selection for genomic DNA fragments in order to enrich for the CG-rich parts of a genome before next-generation sequencing. However, the size fraction taken during the selection step of standard RRBS protocols seemed somewhat arbitrarily chosen and allows for only a very small portion (often just around 1%) of a typical genome to be included in the analysis, with little or no ability to assure that given regions of interest, such as certain genes or regulatory regions are covered. üRRBS-Predictor is centered around an algorithm that will analyze the whole genome and report back where in the genome MspI will cut (fragment start and stop positions), the fragment lengths and the number of CpGs found in each fragment. The user can then explore which portion and percentage of a genome would be covered by RRBS sequence data if a given size fraction of DNA fragments is selected during RRBS sample preparation (in 50 bp increments between 50 and 800 bp), based on experimental limitations). At first this can be done for the human genome and the genomes of the most common model organisms such as mouse, rat, fruit fly, zebrafish, C. Elegans (later-on additional genomes can be included in the webtool). The script produces a BED file that can be uploaded to the UCSC Genome Browser tool that will display the fragments in their locations and color-coded based on the size of the fragment. Alternatively, the user will be able to select a genomic region of interest and will be given all the fragments that cover that region along with their sizes. üRRBS-Predictor is for a computer, for a user, could be enhanced-RRBS-predictor/eRRBS-predictor, variable-RRBS/RRBS - or something entirely different].
1516T

Analysis of archived residual newborn screening bloodspots after whole genome amplification, using Genfromatic's genomic medicine tool suite. D. Weaver1, B. Cantarel1, J. Reese1, R. Finnes2, 3 1) Genfromatic, Austin, TX; 2) University of Texas, Austin; 3) Dell Pediatric Research Institute.

A wealth of genomic health information is archived in residual newborn blood samples originally collected and used for routine newborn screening for common genetic diseases. Where legally and ethically permissible, exome or whole genome analysis of archived samples could potentially reveal valuable epidemiological, population genetic and personal genomic information. We have utilized archived DNA samples containing data from newborn bloodspots obtained from biobanks to test the assumption that archived newborn blood samples are a vast resource for genetic research. By using Genfromatic's cloud-based technology to analyze archived blood samples, we have uncovered valuable information on genetic conditions that may help benefit newborns who have tested positive for these conditions in the past.

1518W

Massively parallel sequencing as a tool for HLA typing. W. Yang1,2, Y. Huang1,2, J. Yang1,2, P.C. Sham2, Y.L. Lau1. 1) Paediatrics & Adolescent Med. Univ Hong Kong, Hong Kong, 10000, Hong Kong; 2) Centre for Genomic Sciences, the University of Hong Kong.

It has been shown that massively parallel sequencing (MPS) for HLA typing has the potential to become a more accurate, high throughput, and highly flexible HLA typing platform for both clinical laboratories and research settings. We introduce the two areas we have made progress in using MPS for HLA typing, including improvement in using genome-wide sequencing (GWS) data to make HLA calls and pooling of a large number of samples for HLA typing using MPS technology. HLA typing from GWS data using available tools has many issues, often because they don't fully account for the polymorphic nature of this group of genes and enormous allelic differences. Thus we developed a multiple references-based mapping approach for defining HLA types from GWS and identified large number of previously unmapped sequencing reads, which allowed a much more complete HLA calling from GWS data. A toolkit was developed to facilitate using this method for HLA typing from GWS. Making use of the available whole exome sequencing data we have and using this novel HLA mapping and calling method, we were able to gather detailed information on the sequences of the polymorphic exons of the major HLA genes and to design PCR primers for amplifying these exons. The primers are particularly suited for East Asian populations. In research laboratories, we often need to type certain HLA genes on a large number of samples, which is quite different from the practices in clinical laboratories. This is especially important for genetic studies of autoimmune diseases and population screening for certain HLA gene alleles to avoid adverse drug responses. Making use of the method and primers we have developed, together with an innovative barcoding system to allow pooling of up to a thousand PCR amplicons together for next generation MPS sequencing, we were able to increase throughput of HLA typing and reduce sequencing cost significantly, while providing flexibility and scalability to the MPS technology.

1517F


With the abundance of information and analysis results being collected for genetic loci, user-friendly and flexible data visualization approaches can inform and improve the analysis and dissemination of these data. An ideogram is a graphic representation of chromosomes, and these plots have been used with the addition of overlaid points, lines, and/or shapes, to provide summary information of various genomic and phenotypic data. We offer a flexible software tool PhenoGram, which exists as a web-based tool and also a desktop application. PhenoGram is easy to learn, and can be used to visualize large amounts of genomic information. It is suited for East Asian populations.

1519T


Accurately calling single nucleotide polymorphisms (SNPs), insertions and deletions (INDELS) is crucial to medical and population genetics. A variety of algorithms have been developed to detect SNPs and have been successfully applied to large-scale re-sequencing projects, but detecting INDELS still presents significant problems. Moreover, most of the existing algorithms are designed for high-quality short-read data with a low INDEL sequencing error rate whereas upcoming technologies such as ion Torrent sequencing technique produces long reads at the cost of an increased sequencing error rate. The high time complexity of the PHMM model has many issues, often because they don't fully account for the polymorphic nature of this group of genes and enormous allelic differences. Thus we developed a multiple references-based mapping approach for defining HLA types from GWS and identified large number of previously unmapped sequencing reads, which allowed a much more complete HLA calling from GWS data. A toolkit was developed to facilitate using this method for HLA typing from GWS. Making use of the available whole exome sequencing data we have and using this novel HLA mapping and calling method, we were able to gather detailed information on the sequences of the polymorphic exons of the major HLA genes and to design PCR primers for amplifying these exons. The primers are particularly suited for East Asian populations. In research laboratories, we often need to type certain HLA genes on a large number of samples, which is quite different from the practices in clinical laboratories. This is especially important for genetic studies of autoimmune diseases and population screening for certain HLA gene alleles to avoid adverse drug responses. Making use of the method and primers we have developed, together with an innovative barcoding system to allow pooling of up to a thousand PCR amplicons together for next generation MPS sequencing, we were able to increase throughput of HLA typing and reduce sequencing cost significantly, while providing flexibility and scalability to the MPS technology.
1520F Resolving complex structural genomic rearrangements using a randomised approach. X. Zhao1, S.B. Emery2, J.M. Kidd1, R.E. Mills1,2. 1) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Genetics, University of Michigan, Ann Arbor, MI.

Structural variants (SVs), defined as the deletion, duplication, insertion, inversion or translocation of genomic regions, are both a major source of genetic diversity in human populations and are also directly responsible for the pathogenesis of numerous diseases. Many studies have been conducted in the past decade to discover and analyze SVs, however these have predominantly focused on unbalanced (copy number variant) events involving only one or two breakpoints. In contrast, more complex rearrangements resulting from multi-step or overlapping events involving three or more breakpoints have received considerably less attention or have been incorrectly interpreted. We have developed an algorithm, metroSV, which accurately analyzes and resolves complex rearrangements using whole genome, paired-end sequencing data. This method first identifies regions of the genome suspected to involve a complex event and then delineates putative breakpoints using aberrant sequence alignments. The resulting segments are then iteratively rearranged in a randomized fashion and scored against expected models of sequence characteristics using a Markov chain Monte Carlo approach to infer the underlying architecture of these variants. We have applied our algorithm to well-characterized genomes including the high coverage Illumina Platinum CEPH pedigree that has been deep sequenced and compared our results to previously reported complex events in these samples. We are in the process of broadening the scale of our algorithm to assess the entire genome concurrently and are also expanding our analysis to other large-scale, publicly available datasets as well as pathogenic samples associated with germline or somatic chromothripsis-like events. We believe metroSV represents a significant advancement towards resolving these complex chromosomal structural rearrangements and furthering our understanding of their mechanistic origins and functional impact.

1521W Systematic integration of functional and computational genomics suggests that the indel rs79240969 in the DN33 gene influences both bone- and obesity-related traits. M. Clausnitzer1, X. Chen1, D. Karasik1, LA. Cupples1, DP. Kiel1, YH. Hsu1. 1) Inst Aging Res, Hebrew Senior Life, Harvard Medical School, Boston, MA; 2) Boston Univ Sch Pub Hlth, Boston, MA.

Recent studies suggest shared etiologies of bone and adiposity phenotypes. Previously, we conducted a multivariate genome-wide association study (GWAS) of bone mineral density (BMD) at different skeletal sites and metabolic syndrome risk factors, and identified 31 bivariate genome-wide significant loci (p<5x10-8, p(bivariate)<p(univariate)/10). However, signals emerging from GWAS are merely markers for large genomic regions in linkage disequilibrium (LD), harboring the disease-causing variant. Identifying the shared causal variants is central to elucidate the molecular mechanisms underlying the genetic correlation between bone and fat. We applied an integrative bioinformatics approach leveraging tissue-specific functional and computational genomics and sequence data to narrow-down the potential causal variants. We identified all reported sequence variants (from the 1000 Genomes project) within the identified bivariate GWAS loci (physical boundaries LD r2>0.7). We merged data on sequence variants with bone- and adipose-specific genome-wide epigenomic profiling data, reported from the ENCODE project, that allow for chromatin state-dependent analyses of regulatory variation. Within predicted regulatory regions, we discovered cis-regulatory modules (CRMs) by analyzing cross-species conserved patterns of transcription factor binding sites (TFBS) across 16 vertebrate species. Using our approach, we identified potential causal variants within the bivariate association loci that may be responsible for the association with BMD and obesity traits. One example is the bivariate signal at DN33, 1q24.3 (GWAS SNP rs10489290, bivariate association test p=8.4 x 10-11 for femur neck BMD with waist-to-hip ratio). We pinpointed an intronic insertion/deletion variant rs79240969 (−/TCA, MAF=0.296, LD r2=1.0), specifically mapping within predicted osteoblast and adipocyte gene regulatory regions. The cell-type specific TFBS pattern analysis revealed rs79240969 localizing within a cross-species conserved CRM relevant to bone and fat. The rs79240969 insertion allele creates a perfect binding site for the zinc-finger protein Zfp521 which has been previously shown to control bone mass. Our bioinformatics analysis may represent a useful step toward pinpointing causal variants from potentially pleiotropic loci for direct laboratory validation and ultimately for improving therapeutic strategies addressing shared etiological mechanisms of BMD and obesity-related phenotypes.

1522T The Developmental Brain Disorders Database (DBDB): A curated neurogenetics knowledge base with clinical and research applications. G. Mirzaa1, AJ. Barkovich2, WB. Dobyns2, KJ. Miller1, R.E. Piaciorkowski2. 1) Department of Human Genetics, Seattle Children’s Hospital, Seattle, WA; 2) Department of Pediatrics, University of California San Francisco, San Francisco, California; 3) Depts of Neurology, Pediatrics & Biomedical Genetics, Center for Neural Development & Disease, University of Rochester Medical Center, Rochester, NY.

The number of genes associated with neurodevelopmental disorders has risen dramatically over the past decade, with the discovery of many genes associated with intellectual disability, autism, epilepsy, and others. The maturation of copy number studies and the use of whole exome sequencing (WXS) have further increased the number of genes associated with neurodevelopmental phenotypes. For geneticists, this rapid expansion of knowledge increases the difficulty of arriving at diagnoses in a timely and cost-effective manner. Given rapidly advancing technology, the genetics community has moved most of its molecular and clinical resources online to assure more rapid dissemination of information. The currently available resources, while encyclopedic, are limited by the state of organization of the knowledge base. There is no system of levels of evidence for gene-phenotype associations, making it difficult to judge which genes ought to be tested first. Also lacking is a neurodevelopmental phenotype ontology. Finally, laboratories working with WXS data would benefit from a well-curated source of genes associated with disorders of brain development. To address these issues, we designed and implemented a publicly available web-based tool that curates the body of knowledge regarding genes associated with neurodevelopmental phenotypes, assembled the first ontology of those phenotypes, and developed a set of tools for comprehensive analyses. This tool is called the Developmental Brain Disorders Database (DBDB), an on-line curated repository of genes, phenotypes, and syndromes associated with neurodevelopmental disorders available at https://www.dbdb.umr.ochestcr.edu/home. The current release contains over 700 evidence-ranked gene-phenotype associations, 70 neurodevelopmental phenotypes, and 150 syndromes. While DBDB augments existing web-based resources such as OMIM and GeneReviews, it uniquely offers an online evidence system for knowledge of neurodevelopmental disorders that is integrated into the literature, and curated by recognized experts. As new literature emerges, DBDB is easily updated. References remain current as they are served directly from PubMed. When used with other resources, DBDB will streamline the genetic workup of children with neurodevelopmental disorders.

For researchers, it provides an evidence-ranked gene list against which results of WXS may be filtered, and facilitates more advanced pathway-based analyses.
1523F

Identifying putative functional variants from GWAS by utilizing ENCODE Consortium data. J. Hayes1,2,3, X. Xu1,2,3, J. Faber1,2, M. Scherf1,2, A. Gonzalez1, A. Perez1,2, V. Joseph1,2, K. Off1,2, S. Raychaudhuri1,2, C. Leslie1, R. Klein1,2. 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 2) Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 3) Biochemistry, Cellular, and Molecular Biology Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA; 4) Immunology and Microbial Pathogenesis Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA; 5) Physiology, Biophysics and Systems Biology Program. Weill Cornell Graduate School of Medical Sciences, New York, NY, USA; 6) Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 7) Tri-institutional MD-PhD Program, Weill Cornell Medical College, New York, NY, USA; 8) Divisions of Genetics & Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 9) Partners Healthcare Center for Personalized Genetic Medicine, Harvard Medical School, Boston, MA, USA.

The first genome-wide association study (GWAS) was published in 2005. Since then, there have been over 1,500 published GWAS papers spanning over 200 diseases and more than 3,000 disease susceptibility loci identified. However, one of the major challenges of the field is that over 90% of these identified risk variants do not result in an amino acid change, thus failing to explain the mechanism for increased risk. We hypothesize that the GWAS-identified risk-SNPs, or those variants highly correlated with them, increase risk is by altering transcriptional regulatory elements. We tested this hypothesis by analyzing identified risk-SNPs for both lymphoma and chronic lymphocytic leukemia (CLL) together. The lymphoblastoid cell lines (LCLs) in ENCODE make for an ideal dataset in which to test these risk SNPs for B-cell malignancies as LCLs are derived from B cells. When combined, there are 29 non-HLA lymphoma risk SNPs published to date. When we identify tagged SNPs that are in LD with the original SNP with an r2 > 0.8 the number of SNPs to be analyzed increases to 352. When we compare these with our DNase hypersensitivity data, 23 SNPs lie in hypersensitivity regions. We then overlapped these DNase-SNPs with ChromHMM data and found 11 SNPs that were DNase rich. These SNPs are potential functional SNPs and provide a framework by which other groups can accurately and quickly select these SNPs for preliminary experimental follow-up.

1524W

Prioritization of Pathways, Genes and Polymorphisms for Dental Caries. A. Nicolaou1,2,3, A. Gonzalez1, A. Perez1,2, V. Joseph1,2, K. Off1,2, S. Raychaudhuri1,2, C. Leslie1, R. Klein1,2. 1) Center for Pharmacogenomics and Complex Disease Research, New Jersey Dental School, UMDNJ, Newark, NJ; 2) School of Health Related Professions, UMDNJ, Newark, NJ.

The main goal of this study was to identify genetic variation that may explain why some individuals are more susceptible than others to dental caries (tooth decay or ‘cavities’). The discovery of causal genes and awareness of risk factors is essential for development of personalized prevention and early detection of this major oral disease. Although oral bacteria play an important role in the development of dental caries, there is strong evidence for genetic risk of dental caries from human twin and family studies. Our study used two sources for selection of genes and Single Nucleotide Polymorphisms (SNPs): 1) the biological roles of genes as reported in Gene Ontology (GO) and 2) genetic association studies reported in literature. We found that GO was limited in the coverage of oral biology. However, good candidate genes were found for enamel, antimicrobial and saliva pathways. After the first Genome Wide Association Study (GWAS) for caries and the candidate genes obtained in our study may only weak to moderate statistical support. Therefore, it appears likely that the selected genes/SNPs from our study and those obtained from the GWAS. These developments risk demoting the clinical team to bystanders, and could disrupt physician engagement. Approach/Results. We developed the human Genomic Clinical Annotation Tool (h-GCAT), a server-based software tool that allows the user to upload a variant call format file (vcf), containing the variants of patient(s) and relevant family members. The tool allows quick selection of relevant cut-offs for sample sequencing depth and genotype quality, minor allele frequencies (using 1kG and ESP5600), and of relevant chromosomes. Subsequently, the user continues to filter using pedigree information and type(s) of mutation(s) to be ascertained (known/Novel SNPs, gene/exon/coding region, synonymous/non-synonymous SNPs). The user can also filter by clinical annotation, using tools as HGMDB, DO, OMIM and HPO. Regions of homozygosity can be analyzed in cases with consanguinity. At the end of this process, the user can review the calls that ‘survived’ these iterative filters. The result page provides relevant information of these calls: genomic location of the candidate calls, dbSNP status, cDNA location of calls for the various transcripts, as well as codon change by protein alignment(s), if applicable. Link Outs to the NCBI, UCSC Genome Browser, OMIM, HGMDB and DO entries are provided for review. We tested our tool on various training sets given certain inheritance patterns. Within less than 10 minutes results similar to other WES analysis tools were obtained. Conclusion. While the software underlying the various algorithms of filtering activities in h-GCAT is intricate, the process of filtering out (likely irrelevant calls and performing a directed search for relevant mutation(s) and associated phenotypes is easily understood and accomplished by a clinician with little bioinformatics training. We found that current software tools that can inform phenotyping, becoming more important with continued identification of Mendelian disorders. It is our opinion that variant call files should be made available to the clinician when requested.

1526F


Association studies based on next-generation sequencing (NGS) technology have become popular, and statistical association tests for NGS data have been developed rapidly. A flexible tool for simulating sequence data in either unrelated case-control or family samples with different disease and quantitative trait models would be useful for evaluating the statistical power for planning a study design and for comparing power among statistical methods based on NGS data. To our knowledge, the software SimRare is the only tool designed specifically to simulate sequence data with phenotypes. However, SimRare focused on generating unrelated samples. As family-based association studies using NGS are also important, software that can simulate sequence data in families will be very useful for evaluating the properties of family-based NGS analysis. We developed a simulation tool, SeqSIMLA, which can simulate sequence data with user-specified disease and quantitative trait models. We implemented two disease models, Models 1 and 2, based on logistic functions. In Model 1, the user can flexibly specify the odds ratios of disease variants and disease prevalence. In Model 2, the user can specify the heritability and population attributable risk for disease variants and the baseline penetrance. Risk and protective variants can also be simulated. We do not have restrictions on the number of disease loci to be simulated. We also implemented a quantitative trait model, in which the user can specify the number of quantitative trait loci (QTL), proportions of variance explained by the QTL, and genetic models. We compiled recombination rates from the HapMap project so that genomic structures similar to the real data can be simulated. Both unrelated case-control and three-generation family data can be simulated. Java threads are used to parallelize the software underlying the various algorithms of filtering activities in h-GCAT. SIMLA will be very useful for evaluating statistical properties for new study designs and new statistical methods using NGS. SeqSIMLA can be downloaded at http://seqsimla.sourceforge.net.
1527W
RegScan: a tool for rapid estimation of allele effects on continuous traits and their combinations. T. Haller1, M. Kale1, T. Esko2,3, R. Mägi1, K. Fischer1. 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Broad Institute, Cambridge, MA, USA; 3) Children's Hospital Boston.

Genome-wide association studies (GWAS) are becoming computationally more difficult with the ever-increasing number of samples, markers, and traits. The combinatorial traits (for example the ratios of metabolite concentrations) can increase the data size beyond what the current GWAS tools can handle. At the same time the scientific interest to study combinatorial traits is rapidly increasing and the field needs a quick analytical tool. We addressed the issue of increasing need for computational power in GWAS by creating an application called RegScan. Our tool is designed for performing basic linear regression analysis with very large data sets maximally fast. It performs association analysis between markers and continuous combinatorial traits ten to hundreds of times faster than the other tools. RegScan specifically targets association analysis of combinatorial traits in metabolomics, however, it is not limited to that. Another area of use is studying marker associations with gene expression. RegScan can automatically create the combinatorial traits for analysis. It also comes with several supporting functions used for filtering of the results and additional analysis of the detected associations with the focus to determine their significance. RegScan has been successfully tested in linear regression analyses of large data sets such as 1000 Genome imputed data for thousands of individuals with over 6000 combinatorial traits. The analyses that have typically taken several weeks or months can now be carried out in just days. The results allow researchers to investigate the links between genetic markers and biological pathways. RegScan is an open source project. It can be freely downloaded together with the instructions and examples at www.biobank.ee/regscan.

1528T
Shortening the Diagnostic Odyssey: Integrating Genomic, Structural, and Phenotypic Information to Reduce Time of Rare Disease Diagnosis. R. Hantharan1, M. Bhat1, S. Agarwal1, S. Krishna1, B. Panda1, V. Veeramachaneni1. 1) Strand Life Sciences, Bangalore, India; 2) Center for Human Genetics, Bangalore, India.

Genomic sequencing for diagnosis of rare, and often complex, diseases is gaining traction to help end what is often a diagnostic odyssey for families and patients. However, knowledge of genetic variants alone is often not enough to diagnose disease - understanding the effects of variants on gene and protein structure inform how the variant may affect the role in biological function while incorporating phenotypes and symptoms help explain how the variants manifest in patients. To that end, we are developing software that combines our own tool for analysis of NGS data (AVADIS NGS) with systematically curated literature content and bioinformatics databases (to integrate genomic, structural and phenotypic information) to enable data to report generation in one single step. This will substantially compress the time needed for clinicians to correctly diagnose patients and start them on the right treatment regimen. Here, we present a case study wherein, through collaboration with a clinical geneticist, we have sequenced and applied our approach to a family in India. The young couple, of a consanguineous marriage, had previously given birth to two boys with pulmonary hypertension and respiratory problems who died suddenly under two years of age. The couple was again pregnant and wanted to determine whether this child would experience similar outcomes. Through analysis of whole exome sequencing data of the parents, fetus, and one of the deceased brothers with our software, we were able to provide an answer.

1529F
Gene Expression Deconvolution using Single-cells. J. Lindsay4, I. Mandoli4, C. Nielson4. 1) Department of Computer Science and Engineering, University of Connecticut, Storrs CT; 2) Department of Molecular and Cell Biology, University of Connecticut, Storrs CT.

Obtaining whole-transcriptome expression profiles of closely related cell types is challenging for stem-cell biologists. Here we present an approach that utilizes single-cell qPCR probing of a small number of genes to aid in the deconvolution of whole-transcriptome profiles of mixed samples. Typically the expression profiles of a given mixture of cells is modelled as linear combination of the signature of its constituent cells multiplied by the concentration of each cell type in the mixture. Existing approaches to deconvolution methods attempt to estimate both the cell type signatures and concentrations simultaneously, or separately if knowledge of one is known beforehand. Our method first obtains a reduced profile of constituent cell-types from single-cell samples by using k-means clustering and then averaging all cell-types in each cluster. Then we apply a robust quadratic programming method to inferring mixture proportions of mixed sample. Finally we have implemented a second quadratic program for inferring cell-type specific expression levels of genes not measured directly in single-cells based on mixture proportions derived for each mixed sample. Using real single-cell data obtained from the posterior Node-Streak-Border region of a mouse embryo and 100 simulated mixtures, a leave-one-gene-out experiment found our method estimates of concentrations had a RMSE of 0.03 and the missing gene estimates had a correlation of 0.997.

1530W
Implementing a High Performance, Reusable Consensus Calling Pipeline for Next Generation Sequencing using Globus Genomics. R.K. Madduri1, A. Rodriguez1, V. Trubetsky2, L.R. Davis2, P.J. Dave1, N.J. Cox2, I.T. Foster1. 1) Computation Institute, University of Chicago, Chicago, IL; 2) Section Genetic Medicine, University of Chicago, Chicago, IL.

We developed Globus Genomics (http://globus.org/genomics/), an end-to-end hosted service designed to efficiently and easily analyze large quantities of Next Generation Sequencing (NGS) data using state-of-the-art algorithms, efficient data management tools, a graphical web-based workflow environment and on-demand computing infrastructure. Globus Genomics leverages a collection of existing cloud-based services. Globus Genomics users, however, can build new analysis workflows from scratch. Users can analyze large amounts of data using computationally efficient analytical pipelines and cutting edge tools that leverage the power and flexibility of on-demand cloud computing resources—without being exposed to the complexities of managing large scale infrastructure; deploying and configuring analysis tools; transferring data between sequencers, analysis nodes and storage systems; or managing their own users and groups. To this end, we use elastic computational infrastructure provided by Amazon Web Services. We use the Condor scheduler to manage a dynamically assembled pool of hosts. We outsource high performance data transfer and user, group and credential management to Globus Online, a platform as a service (PaaS) provider also developed and operated by our team. Finally, we host a Galaxy workflow system to enable easy to use graphical workflow orchestration. We created computational pipelines for multiple variant calling and genotyping algorithms available for academic use (i.e., GATK2.0, Atlas2.0, and FreeBayes toolkits). These profiles enable high performance, scalable execution of algorithms on hundreds of raw data sets. We built reusable, robust pipelines using different computational modalities that best suited the underlying analysis. The resulting variant calls from each pipeline can then be fed to a consensus-calling algorithm (Consensus Genotype for Exome Sequencing CGES; see Trubetsky et al., ASHG 2013) resulting in high quality variant and genotype calls. We have run these three pipelines in parallel calling variants on over a hundred raw BAM files in the course of three days. Atlas2.0 and the GATK pipelines took a little over two days to finish execution while Freebayes pipeline took a little over three days. In conclusion, we present the workings of Globus Genomics, a robust, powerful, and user-friendly suite of tools for NGS analysis empowering geneticists and enabling translational discovery relevant to human disease.
1531T Genome-wide structural variation analysis with genome mapping on nanochannel arrays. A.C.Y. Mak1, J.J.K. Wu2, Y.Y.Y. Lai1, K.Y. Yip3,5, T.F. Chan1,5, E.T. Lam1, T.P. Kwok4, J.W. Li6, A.K.Y. Leung4, A.K.Y. Yim4,5, M. Xiao6, P.Y. Kwok6, S.M. Yu6, 1) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 2) Department of Computer Science, The University of Hong Kong, Hong Kong; 3) Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong; 4) School of Life Sciences, The Chinese University of Hong Kong, Hong Kong; 5) Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Hong Kong; 6) School of Biomedical Engineering, Science & Health Systems, Drexel University, Philadelphia, PA.

Despite recent advances in next-generation sequencing technology, genome-wide structural variation (SV) detection using ‘short reads’ remains challenging. Detection of large and/or balanced SV such as inversion or translocations is difficult, if not impossible. To overcome the limitations of short reads, we generated genome maps using a novel approach that allows very long DNA molecules (> 150kb) fluorescently labeled at Nt.BspQI sites (GCTTCTCN) to be linearized and imaged in highly parallel nanochannel arrays [1]. We obtained data to 50X genome coverage on NA12878, a member from a CEPH CEU trio extensively sequenced in the 1000 Genomes Project. To detect structural variations from these genome maps, we have developed a BreakPoint sensitive Optical Map Dynamic Programming algorithm (BP-OMDP). Data for each DNA molecule is represented as an array of inter-label distances or segments. BP-OMDP allows molecules to be partially aligned to the hg19 in silico Nt.BspQI reference with mismatching information for each SNP reside allows random access to any marker. Average compression ratios of ~90% are observed in test data.


Genome-wide association studies directly assay ~10^6 single nucleotide polymorphisms (SNPs) in a study population. With access to phased reference samples of common ancestry that have genotype information at a denser panel of variants, a sample of 10^5 SNPs can be extended probabilistically by a factor of 10- to 40-fold. These resulting datasets are standardly stored as text files that detail imputed genotype probabilities, requiring 12 characters per SNP per sample, thus are too large to be simultaneously stored in RAM losslessly as is. Instead, the maximum likelihood genotype at each site is used, or SNPs are processed and discarded one at a time. Neither of these adaptations are optimal: missingness at individual sites may be informative to a trait, and multi-locus testing is complicated. Here, I present a C/C++ library that dynamically compresses probabilistic genotype data as they are loaded into memory. This method uses a customization of the DEFALATE (gzip) algorithm, and maintains constant-time access to any SNP (but linear O(# individuals) access to a given genotype). Briefly, the probabilistic genotype matrix is stored one SNP at a time, with each SNP data being compressed. Genotype calls for individuals for a particular SNP require serial access, but indexing the compressed matrix for where does information for each SNP reside allows random access to any marker. Average compression ratios of ~90% are observed in test data.

1533W VariantMaster: a novel platform to identify causative variants from HTS data in familial, denovo and somatic genetic disorders including cancer. P.A. Santoni1, P. Makryhydramas1, S. Nikolaev1, M. Giupponti2, D. Roby3, A. Bottin4, S. Antonarakis1,2,4, 1) Department of Medical Genetics and Development, University of Geneva, Geneva, Switzerland; 2) Geneva University Hospital - HUG; 3) iGE3 institute of Genetics and Genomics of Geneva, Switzerland.

There is a fast growing interest in clinical genetics to the employment of High Throughput Sequencing data for accurate diagnosis of monogenic diseases. Furthermore, whole exome sequencing data introduced a significant role of changes in the complexity of cancer development by identifying the identification of driver somatic variants. To the aim of providing an accurate and efficient methodology in these contexts, we developed VariantMaster, an original and comprehensive methodology that extracts likely causative variants in familial and sporadic genetic diseases considering different modes of inheritance: X-linked, autosomal dominant, and recessive (homozygosity or compound heterozygosity), denovo germline and somatic mutations including cancer. For the highest accuracy the algorithm takes into account predicted variants and integrates the phenotypes with family pedigrees. A probabilistic approach based on row data is furthermore applied to robustly evaluate the likelihood of the occurrence of a putative causal variant in each family member. Additionally, VariantMaster can incorporate several layers of information as, for example, genotypes, allelic frequencies and damaging scores. VariantMaster is already employed as an effective tool in the clinical genetics department of the Geneva University Hospital (HUG) and in several research projects as the identification of the burden of denovo mutations in a cohort of 58 trios to elucidate the molecular basis of schizophrenia and in the identification of novel causative genes in individuals with Callosal agenesis (PCD). Furthermore, to prove the effectiveness of the methodology on the detection of somatic variants we extract 9 random samples of colorectal cancer tumors from the TCGA consortium. Remarkably, 80% of the variants identified by VariantMaster were validated and annotated in COSMIC. Moreover, VariantMaster detected two novel putative ‘driver’ variants. In general VariantMaster demonstrated to be more flexible and accurate compared with previously published algorithms and we believe it has the potential to become an indispensable tool in the investigation of genetic diseases and molecular cancer profiles.


Purpose: STAT3 is activated in response to external signals. After tyrosine phosphorylation of the receptor, STAT3 is recruited by its SH2 domains, is phosphorylated, dimerizes and translocates to the nuclease and binds to specific promoter sequences. Many dominant-negative STAT3 mutations have been identified from cases of hyper IgE syndrome, and activating mutations from large granular lymphocytic leukemia (PCD). In the protein dimerization interface of the SH2 domain from leukemia patients, we examined the ability of 9 computational methods to predict the effect of these mutations on protein function. Methods: Nonsynonymous mutations were initially detected by large scale sequencing of chronic lymphocytic leukemia patients and were subsequently annotated in COSMIC. The majority reported in multiple patients each) and 10 mutations in the (PCD). In this group, 80% of the variants identified by VariantMaster were validated and annotated in COSMIC. Furthermore, VariantMaster detected two novel putative ‘driver’ variants. In general VariantMaster demonstrated to be more flexible and accurate compared with previously published algorithms and we believe it has the potential to become an indispensable tool in the investigation of genetic diseases and molecular cancer profiles.
1535F
A Graphical Quality Control Tool for Next Generation Exome Sequencing. J.D. Smith, S. McGee, D.A. Nickerson. Genome Sciences, University of Washington, Seattle, WA.

The wide-scale application of massively parallel sequencing requires the development of new approaches to automate data quality control. These approaches should accurately and efficiently pinpoint potential problems and evaluate quality throughout the process. Additionally, the quality metrics associated with the data need to be presented in a concise but informative format to facilitate rapid decision-making about whether to proceed to interpretation, or to remove or hold a sample for further assessment. To facilitate this process, we have developed a “QC Dashboard” that presents an overview of quality metrics, and is currently in use for exome sequencing at the Northwest Genomics Center (NWGC). The QC Dashboard is a graphical and textual display of statistical information extracted from BAM files, and uses multiple open-source analysis tools. Plots include overall number of reads, reads mapped to target and number of unique reads; allele distributions at each sequencing cycle; plots for read depth coverage, uniformity and insert size distribution. Per sequencing cycle error rates and read qualities are also plotted and summarized. Sample complexity is quickly and accurately extrapolated from a subset of data and displayed. Each dashboard is adaptable to individual- or merged-lane data and can easily be implemented in any analysis pipeline for automatic display.

1536W
Evaluation of imputation method for classical HLA-DRB1 using a Finnish dataset. E. Viachopoulos1, E. Lahtela1, A. Wennerström1, A.S. Havulinna2, P. Salo1,4, M. Perola2, V. Salomaa1, M.S. Nieminen1, J. Sinisalo1, M.L. Lokki1. 1) Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland; 2) HUCH Heart and Lung Center, Division of Cardiology, Helsinki University Central Hospital, Finland; 3) National Institute for Health and Welfare, Helsinki, Finland; 4) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland.

HLA genes are located at the MHC region (6p21) revealing high degree of gene density and polymorphism. Allelic structures of HLA genes show broad linkage disequilibrium and are inherited as haplotypes with significantly different frequencies between populations and ethnicities. HLA genes have been associated with many autoimmune, infectious and inflammatory diseases. Due to vast amount of alleles, the high-resolution HLA typing is expensive and time-consuming. Scientists have attempted to develop computational approaches to define HLA alleles with high confidence. We tested the reliability and performance of a widely used software HLA1MP (versions HLA1MP-01 and HLA1MP-02) for imputing classical HLA-DRB1 alleles in the Finnish material. Unrelated samples (n=161) from Finnish subjects were selected from the FINRISK 1997, 2002 and 2007 cohorts. The samples were SNP genotyped with Illumina 610K genotyping chip and used as the input for HLA1MP. The frequencies of the imputed HLA-DRB1 alleles were compared with the high-resolution results obtained from sequence-based typing. The frequencies of the imputed HLA-DRB1 alleles were similar to the frequencies from sequence-based typing. However, the per-individual success rate was 27.64% using either HLA1MP-01 or HLA1MP-02 having mean Q and Q2 posterior probabilities for correct inferences of 0.9329 and 0.9286, respectively. When we used the posterior probability of 0.95 as a threshold for both imputed alleles simultaneously, the per-individual success rates were as low as 23.74% and 23.39% with HLA1MP-01 and HLA1MP-02, respectively. Hence, the higher confidence decreased the correct inferences. One of the most prominent example was HLA-DRB1*01:01 allele showing approximately 30% success rate while being the most common wrongly imputed allele. In Finland, isolation and migration history have shaped the gene pool narrower showing HLA haplotype frequencies typical to Finnish population when compared to Europeans. When we used HLA1MP having different European populations as references, the imputation success for HLA-DRB1 alleles was very low pointing to the importance of population specific reference material.

1537T
An adaptive permutation procedure to estimate the significance threshold for the minimum p-value of multiple permutation tests. P. Yajnik, M. Boekhoorn, H. J. van Dongen, Dept Biostatistics, University of Michigan, Ann Arbor, MI.

The analysis of modern genetic association studies often involves testing multiple hypotheses. The Bonferroni correction is commonly used to approximate p-value significance thresholds to maintain the familywise type I error rate. However, when tests are correlated, the Bonferroni correction may be conservative and permutations can be used to estimate significance thresholds. Permutation based estimates are computationally expensive. The computational burden is exacerbated if p-values of the tests also need to be obtained by permutation when analytical values based on asymptotic theories are unavailable/inaccurate. We propose an adaptive permutation procedure which can greatly reduce the total number of permutations needed to estimate the minimum p-value of multiple permutation tests. At each iteration, the pool of tests included in the procedure may be reduced by discarding tests whose estimated p-values are much larger than the currently estimated minimum p-value. The likelihood-ratio test (LRT) statistic is used to discard tests. The proposed estimator is the minimum estimated p-value amongst the tests that remain at this threshold.

We performed 10000 simulations to assess the performance of the procedure. The simulation parameters included the number of multiple tests (100, 200, 500, 1000 or 10000), the correlation between blocks of tests (0, 0.2, 0.4, 0.6, 0.8 or 0.95) and the stringency used to discard tests (LRT statistics exceeding 2.7, 3.8, 6.6 or 10.8). At the highest stringency, the proposed method required on average only 0.2%-5% as many permutations as the naive method (the relative efficiency increasing with number of multiple tests). At this stringency, the proposed method achieved the exact minimum p-value found by the naive method in 99% of the simulated cases.

We also applied this procedure to a real dataset. A burden test (SKAT) was performed with each of 13 correlated phenotypes regressed on variants from 62 genes (806 tests with a Bonferroni corrected threshold of 6e-05). It took our procedure 12 hours with 40 CPUs to provide an estimated threshold of 7e-05. Per our simulations, the naive approach would take 20–500x longer than the proposed approach. Based on a modest estimate of 100x, the naive approach would require 1200 hours.

1538F
BDgene: a genetic database for bipolar disorder and its overlap with schizophrenia and major depressive disorder. S. Chang, L. Guo, J. Wang, Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences. 16 Lincui Road, Chaoyang District, Beijing, China.

Bipolar disorder (BD) is a common psychiatric disorder with complex genetic architecture. It shares overlapping genetic influences with schizophrenia (SZ) and major depressive disorder (MDD). Large numbers of genetic studies of BD and cross-disorder studies between BD and SZ/MDD have accumulated numerous genetic data. There is a growing need to integrate the data to provide a comprehensive data set to facilitate the genetic study of BD and its highly relevant diseases. To fulfill this demand, BDgene database was developed to integrate BD-related genetic factors and shared ones with SZ/MDD from profound literature reading. Through depth-mining of 796 papers, BDgene contains multiple types of literature-reported genetic factors of BD with both positive and negative results, including 797 genes, 3119 SNPs, and 789 regions. Shared genetic factors such as SNPs, genes, and regions from published cross-disorder studies among BD and SZ/MDD were also presented, including 285, 120 and 49 shared genes for BD-SZ, BD-MDD and BD-SZ-MDD respectively. On the basis of data from the literature, in-depth analyses were performed for further understanding of the data: gene prioritization analysis for literature-origin genes obtained 43 BD core genes; pathway-based analysis for genome-wide association study data identified 70 BD candidate pathways; intersection analysis of multidisease candidate genes got 127, 79, and 107 new potential cross-disorder genes for BD-SZ, BD-MDD, and BD-SZMDD respectively; pathway enrichment analysis for BD core genes showed the majority of the pathways were involved in synaptic transmission, membrane and ion channel activity; pathway enrichment analysis for literature-reported positive genes shared by BD-SZ and BD-MDD identified 282 rhythmic related pathways were enriched by both gene lists. As a central genetic database for BD and the cross-first order database for BD and SZ/MDD, BDgene provides not only a comprehensive review of current genetic research but also a powerful source of disease candidate genes and pathways for understanding of BD mechanism and shared etiology among its relevant diseases. To facilitate better usage of the database, BDgene provided powerful search tools to access the data, and a forum to share or exchange ideas. BDgene will be updated regularly for both the literature analysis results to maintain an up-to-date resource. BDgene is freely available at http://bdgene.psych.ac.cn.

Genetic testing, including full gene sequencing and large rearrangement analysis for germline BRCA1 and BRCA2 mutations, is identifying individuals with Hereditary Breast and Ovarian Cancer syndrome. Current genetic analysis identifies BRCA1 and BRCA2 deleterious mutations as well as variants of unknown clinical significance. Reclassification of uncertain variants to more clinically interpretable categories is critical for patient management. We have developed a statistical algorithm that aids in the assignment of clinical classifications to uncertain variants. This algorithm is based on the premise that disease-associated mutations will be observed more often in individuals at high risk for carrying a mutation, as determined by personal and family history. Statistical analysis weights the family histories of each proband carrying the variant of interest and compares these histories to those of control probands carrying variants known to be benign or deleterious. Data from over 400,000 probands were utilized for algorithm development. This technique was validated by and used to analyze over 6000 BRCA1 and BRCA2 variants. The algorithm successfully classified well-documented variants such as BRCA1 c.181T>G (Deleterious), BRCA1 c.1065G>A (Polyphenylalanine), and BRCA2 c.2808_2811del (Deleterious).

The BRCA1 c.5096G>A (Suspected Deleterious with reduced penetrance) and BRCA2 c.7878G>C (Suspected Deleterious with reduced penetrance) mutations were classified as "Not Callable" by the algorithm, consistent with their previous hypomorphic interpretations. This "history weighting" algorithm allows for the accurate reclassification of BRCA1 and BRCA2 uncertain variants and improved clinical management of at-risk patients. The history weighting algorithm has also been successfully applied to reclassification of variants in the Lynch syndrome-associated genes MLH1 and MSH2, and it is currently being modified to allow for analysis of germline variants in MSH6. With additional modifications, this algorithm is expected to be applicable to other autosomal dominant cancer-associated and non-cancer-associated genes.


Tools for exploring and visualizing variation data include the 1000 genomes browser http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes, the 1000 genomes data slicer http://trace.ncbi.nlm.nih.gov/Traces/1kg_slicer/ and the phenotype-genotype data integrator http://www.ncbi.nlm.nih.gov/gap/PheGeni. The clinical remap tool at http://www.ncbi.nlm.nih.gov/gap/remap#tab=rsq will provide sequence coordinates for variations on a clinical RefSeqGene record, and the variation reporter service at http://www.ncbi.nlm.nih.gov/variation/tools/reporter will provide a list of known variants and the functional consequences for a region of interest in BED format or set of variants in HGVS or GVS format. The presentation will show how these services are organized in relation to the primary data archives of GenBank, SRA, GEO, dbSNP, dbVar and dbGaP, and how they can be used as input into external services for research, clinical practice or cloud-based analysis projects.

The database of Genotypes and Phenotypes (dbGaP) is an NIH sponsored repository charged to archive, curate and distribute information produced by studies investigating the interaction of genotype and phenotype. The data submitted to dbGaP includes: individual level molecular and phenotype data; analysis results; medical images; general information about the study; and documents that contextualize phenotypic variables, such as research protocols and questionnaires. The molecular data includes array-based, sequence-based or imputed genotypes; expression, as well as next-generation sequencing (NGS) performed to produce whole exome; whole genome; RNA seq; and epigenomic data. NGS sequence submitted to dbGaP, mostly in the form of BAM files, are processed by NCBI’s Short Read Archive (SRA). Phenotypic data pertaining to study participants or samples consists of any combination of cross-sectional or longitudinal demographic; clinical; laboratory; exposure; or treatment variables. Medical images such as CT; MRI; and retinal scans of the eye are also stored and distributed by dbGaP. Analysis results include summary-level statistical evaluations of the association between the phenotypic variables and the molecular data. All documents, analysis, phenotypic and molecular data are accessioned with stable, unique identifiers that support update of the same clinical study with successive genetic investigations. For example, high-throughput genotyping by one research group, followed by exome sequencing by another group. These identifiers also make it possible to cite the primary data used by published genome association studies in a very specific and stable way. The dbGaP provides unprecedented access to very large genetic and phenotypic datasets, both funded by National Institutes of Health, and other funding agencies worldwide. Public access includes summary data on specific phenotype variables which is linked to study documents; statistical overviews of the genetic information and the position of published associations on the genome. Through controlled access, approved researchers from across the globe may obtain complete statistical analyses, individual phenotype and molecular measures.

This presentation will review new dbGaP features that support high-throughput submission, open access data sets, improved query services and tools to download subsets of data for selected individuals.
1547F

Explore genetic components underlying dental caries through gene-set- and network-assisted approaches. Q. Wang1, P. Jia1, K.T. Cuenco1,2, Z. Zeng1, E. Feingold3,4, L.M. Marazita1,2, L. Wang4,2, Z. Zhao4,5,6,1. 1) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 3) Center for Craniofacial & Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 4) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 5) Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN; 6) Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN; 7) Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN; 8) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN.

Dental caries is a common, chronic, and complex disease leading to a decrease in quality of life worldwide. The impact of genetic factors in dental caries has been recognized for a long time. Here we adopted two strategies attempting to dissect the genetic components underlying dental caries, namely the gene set enrichment and protein-protein interaction (PPI) network analyses. For the former strategy, we applied four complementary gene set enrichment methods to a major dental caries genome-wide association study (GWAS) dataset, which consists of 537 cases and 605 controls. After analyzing 1331 Gene Ontology (GO) terms, we identified 13 significantly associated and 17 marginally associated gene sets, including Sphingolipid metabolic process, ‘Ubiquitin-protein ligase activity.’ ‘Regulation of cytokine secretion,’ and ‘Ceramide metabolic process.’ For the network-assisted strategy, we first prioritized 1214 candidate genes that were collected and curated from four major genetic or genomic approaches (association studies, linkage scans, gene expression analyses, and literature mining) according to the magnitude of evidence related to dental caries. Then we searched for dense modules in the prioritized list of candidate genes by incorporating PPI data, and discovered 23 significant modules that are enriched with genes of great interest. Three major gene clusters were observed among the 23 dense modules, including cytokine network relevant genes, matrix metalloproteinases (MMPs) family, and transforming growth factor-beta (TGF beta) family. Both the GO terms and dense modules we identified encompassed the oral environment related to caries development. Collectively, our gene set enrichment and PPI network analyses provided complementary insights into the molecular mechanisms and polygenic interactions in dental caries.

1548W

Variant Association Tools for association analysis of large scale sequence and exome genotyping data array. G. Wang1, B. Peng2, S.M. Leal1. 1) Baylor College of Medicine, Houston, TX; 2) The University of Texas MD Anderson Cancer Center, Houston, TX.

Currently there is great interest in detecting associations of complex human traits with rare single nucleotide variants (SNVs) using large scale sequence and exome genotyping data. The analysis and quality control for rare variant datasets can be challenging with regards to the analysis of common SNVs in traditional genome wide association studies. Data quality control for sequence data uses information from a variety of difference quality matrices, e.g. genotype quality score, read depth. Statistical association tests for the analysis of rare variants aggregate variants across a genetic region which is usually a gene. Within the gene the variants which are analyzed is usually limited to missense, nonsense and splice site variants; therefore before the commencement of association testing the variants must be annotated. To address the specialized analysis of rare variants we developed variant association tools (VAT), a pipeline that implements best practices for rare variant association studies. Major features of VAT include variant site/call level quality control, summary statistics, phenotype genotype based sample selections, variant annotation, summary statistic generation, and the implementation of rare variant association methods for analysis of qualitative and quantitative traits. The association testing framework implemented in VAT is regression based which readily allows for flexible construction of association models with multiple covariates, weighting (by allele frequencies or predicted functionality), interactions terms, and models for pathway analysis. VAT is capable of rapidly scanning through data using multi-processes computation, adaptive permutation and simultaneously conducting multiple association tests. Results can be viewed as text or graphically. The VAT pipeline is provided to readily facilitate user implementation of novel association methods. The VAT pipeline can be applied to sequence, imputed and genotyping array, e.g. exome chip data. VAT can perform association analyses on small to large scale contemporary studies making use of the latest genotyping and sequencing technologies.

1549T

Variant calling in low-coverage whole genome sequencing of a Native American population sample. C. Bizou1, M. Spiegel1, S. Chasse1, I.R. Sizer1, Y. Li1, E. Ma1, P. Mieczkowski1, J. Sailsbery1, X. Wang1, C.L. Ehlers1, K.C. Wilhelm1,2. 1) Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Psychological Sciences, University of Missouri-Columbia; 4) University of North Carolina High Throughput Sequencing Facility; 5) Department of Molecular and Cellular Neuroscience, The Scripps Research Institute.

For a given sequencing budget, the Low-Coverage Whole Genome Sequencing (LCWGS) strategy allows more samples to be sequenced, and therefore potentially finds a greater number of rare variants. However, the low coverage makes variant calling with standard tools difficult, possibly canceling out any such gains in detection of rare variants. Linkage-disequilibrium (LD) aware variant callers may provide calling rate and accuracy to make LCWGS viable. We examined the performance of LCWGS in 708 whole genome sequences from a population sample of Native Americans. Most of the sequences have coverage between 3X and 12X. We called variants with the LD-aware variant caller Thunder, as well as the single-sample and multi-sample GATK Unified Genotyper. We assessed variant calling through a comparison of the sequencing results to genotypes measured in 641 of the same subjects using a fixed content exome array, which contains variants across the frequency spectrum. The median concordance for Thunder is 97.5%, compared to 65.5% and 90.4% for single-sample and multi-sample calls with the non-LD aware caller. The median improvement of LD-aware calling over the other methods is dependent on depth; for samples with less than 5X coverage, the median improvement is approximately 20%, decreasing to only 5% for samples with greater than 10X coverage. LD-aware variant callers are expected to perform less well for rare variants; for variants at which a single minor allele appears in the data, the multi-sample Unified Genotyper identifies approximately 57% of the variants, while Thunder identifies only 41%. Our results indicate viability of the low-coverage strategy for WGS studies can be achieved through use of LD-aware callers.

1550F

Finding the Clinical Answer in Genomic Sequence: Narrowing the Search Space for Disease-Causing Mutations. S. Garcia1, G. Chandralllake1, M. Clark1, A. Patwardhan1, S. Cheritz1, R. Chen1,2, E. Ashley1,3,4, R. Altman1, J. West1, R. Chen1. 1) Personalis, Inc., Menlo Park, CA; 2) Icahn School of Medicine, Mount Sinai, New York, NY; 3) Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA; 4) Department of Genetics, Stanford University, Stanford, CA; 5) Department of Bioengineering, Stanford University, Stanford, CA.

Diagnosis via genomic sequencing relies on identification of causative disease variants from the tens of thousands of variants present in an exome. Protocols typically apply hard filters to exclude variants by applying expectations concerning which variant types could be disease-causing. Sensitivity suffers when criteria cases that expand the phenotypic or genotypic spectrum of known diseases. Conversely, relaxation of filtering criteria may result in overwhelming numbers of candidates, delaying or preventing diagnosis. We investigated whether a knowledge-based ranking system linking reported clinical information with curated phenotype information combined with ranked genotype expectations based on family history could facilitate the identification of causally related variants without sacrificing the ability to detect novel candidates. Samples with clinical features and causative variants representing a broad range of conditions and variant types (including structural variants of various sizes) were sequenced, and had variants aligned and called. A newly developed database linking clinical features to genes was used to identify and rank candidate genes for each sample. Identification of all possible inheritance patterns, allowing for de novo events and non-penetration, yielded genotype expectations ranked by likelihood. We compared a standard approach using variant pathogenicity predictions combined with simple genotype predictions to our novel approach. Our approach reduced the number of candidate variants requiring manual review by a factor of 25 for autosomal dominant and recessive disorders, and by a factor of 5 for autosomal dominant trait. 597 candidate variants were identified using the standard approach, while our approach ranked 41 candidates. Similarly, in a trio segregating an autosomal recessive trait, our approach ranked four candidates; the standard approach ranked four cases, and had variants associated. The known causal variant was ranked first by our approach. The use of our novel, knowledge-based ranking successfully identifies the most likely causative variants in genomic data, reducing manual review time. With current estimates in the range of 20-60 minutes required for review of each variant, this approach has potential to dramatically improve turnaround time for exome/genome sequencing without sacrificing the potential for novel discovery.
Detecting disease-causing alleles with the human gene connectome. Y. Itan1, J.-L. Casanova1,2,3, L. Abel1,4, S.-Y. Zhang4,5, L. Quintana-Murci1, G. Vogt6, D. Fried6, P. Nitschke6, M. Herman1, A. Abhyankar1. 1) The Rockefeller University, New York, NY; 2) Ncker Hospital for Sick Children, Paris, France; 3) Paris Descartes University, Paris, France; 4) Pasteur Institute, Paris, France; 5) Ben-Gurion University of the Negev, Beer-Sheva, Israel.

Infectious diseases have historically been the greatest killer of mankind. They still account for about 25 percent of all human mortality worldwide. At any time, new epidemics or pandemics can significantly increase this proportion. A small proportion of individuals die from infectious diseases; the collective burden of infection is high because there are many infectious agents. It has become increasingly clear that human genetic background is a key determinant of infectious diseases. To determine the disease-causing allele(s), high-throughput genomic methods are applied and provide thousands of gene variants per patient. However, current methods for validating the true disease-causing single allele at the single patient level are inefficient and extremely time consuming. Moreover, there is no available method for automating the selection of candidate disease-causing alleles at the cohort level, posing a major bottleneck in the field in high-throughput clinical genomics. Resolving this problem will revolutionize the field. We recently developed a novel approach, the ‘human gene connectome’ (HGC) - a concept, method and database that describes the set of all in silico-predicted biologically plausible routes and distances between all pairs of human genes (Itan Y et al. PNAS, 2013). With the HGC, we generated a ‘gene-specific connectome’ for each human gene - the set of all human genes ranked by their predicted biological proximity to the core gene of interest. The HGC was further applied to Mendelian disease studies, by effectively identifying novel herp simplex encephalitis (HSE) morbid alleles in whole exome sequencing (WES) patients data. We hypothesized that within a cohort of patients with the same Mendelian (or nearly Mendelian) disease, the cluster that contains the true disease-causing gene for each patient is the HGC-predicted biologically smallest cluster. Investigating a cohort of 108 exome sequenced HSE patients, we developed and applied a Mendelian clustering algorithm which identifies the biologically smallest HGC-predicted cluster that contains one allele per patient. By that we approximate a solution for an NP-complete algorithmic problem and estimate the disease causing allele for each individual in the cohort.

Incorporating phenotypic information to improve Mendelian disease-gene predictions. A. Javed, S. Agrawal, P.C. Ng. Genome Institute of Singapore, A*STAR, 60 Biopolis Street, Genome, #02-01 Singapore 138672.

We introduce a method which combines phenotypic and genotypic information, within Bayesian framework, in a holistic prediction implicating the gene(s) involved. The patient symptomology is mapped to a list of known disorders using Human Phenotype Ontology. The significance of the match is translated into probability and assigned to the genes implicated for the disorders. The putative role of similar genes is included using random walk with restarts between gene-gene relationships. The HGC was constructed by combining multiple data sources and exhibits strong enrichment of known genetic heterogeneity; pvalue<1e-16 using HGMD blocks of a new study sample to the ones of the same ethnic group in the 1000 genome project. We developed a web tool, Visualization of Haplotype Blocks (VHB), to provide better visualization of the big data with the plots of pairwise linkage disequilibrium (LD). Researchers could upload the variation data of one or several samples in a Variant Call Format (VCF). The statistics of pairwise LD in the selected region are calculated from the variations of study samples and data the 1000 genome project which will be stored in the database of the VHB. Then all variation data of the samples will be parsed and imported into the database schema for temporary storage (30 days). Genome regions and ethnic groups of the 1000 genomes project are selected in the web interface and the results can be displayed in theGeneric Genome Browser locally, or exported as the general feature format (GFF) file, which is the standard output for the genome browser and one type of annotation data format of ANNOVAR software. The VHB web-based tool is available at http://safe.cs.nthu.edu.tw/VHB/. It is implemented in Perl programming language and MySQL relational database, and an Intel Core i7 CPU and 8GB memory computer is used under Linux platform.

Inferring HIV Quasispecies from Deep Sequencing Data. S. Mangul1, N. Wu1, N. Mancuso2, A. Zelikovsky2, R. Sun3, E. Eskin1. 1) University of California, Los Angeles; 2) Georgia State University.

Human immunodeficiency virus (HIV) exhibit high genomic diversity within infected hosts, which may explain resistance to existing drugs. Monitoring and quantifying the HIV population requires inferring set of closely related viral variants, referred to as a quasispecies, and estimating their relatives frequencies. Next-generation sequencing (NGS) is a promising technology for characterizing viral diversity due to its ability to generate large numbers of reads at low cost. However, there is a trade-off among throughput, read length and accuracy for different high throughput sequencing platforms. Error rates in next-generation sequencing platforms, such as Illumina, make it infeasible to use this technology for reconstructing HIV quasispecies and inferring their relative abundances. We suggest to use barcode technique during library preparation allowing to distinguish sequencing error from true viral sequences. Using this technique we are able to obtain high accuracy sequencing reads. Here, we introduce a method for inferring HIV quasispecies from high accurate paired-end reads obtained using the barcoding technique described above. Method consists of the following key steps: (a) build consensus from paired-end reads, (b) count joining information to build conflict graph and assemble quasispecies; (c) infer relative abundances of quasispecies using EM algorithm. To build a consensus from paired-end Illumina reads Vicuna tool is used. Vicuna is de novo population consensus assembly software able to produce single consensus from deep viral sequencing data. We use Mosaic tool with default parameters to map reads to assembled consensus. Based on the read mapping we construct a conflict graph which is used to represent HIV quasispecies in a sample. A conflict graph is a directed graph where each vertex represents unique read. Two reads are connected by an edge if they overlap but disagree on the overlap. To solve the quasispecies assembly problem it necessary to partition the conflict graph into a minimal number of independent sets. This problem also known as graph coloring problem which is NP-hard. To solve this problem we iteratively partition the vertices of the graph two disjoint subsets using greedy algorithm that maximizes cut between this subsets. Preliminary experimental results on synthetic datasets show that our method is able to reconstruct HIV quasispecies from deep sequencing Illumina paired-end data.
155ST Detection of sample-level contamination in next generation sequencing experiments. T.E. Scheetz1,2, A.P. DeLuca1,2, E.M. Stone1,2, T.A. Braun1,2. 1) Dept Ophthalmology, Univ Iowa, Iowa City, IA; 2) Institute for Vision Research, Univ Iowa, Iowa City, IA.

The goal of the this research project is to detect contamination of genomic DNA samples used in next-generation sequencing applications. When performing massively parallel sequencing in a clinical setting, it is critical to ensure that any variations detected result from the correct patient sample - and only from that sample. Known genotype fingerprints can help validate sample identity. But additional quantitative measures are required to ensure sample integrity.

We used the Genome Analysis Toolkit (GATK) to call variations. The relative number of supporting reads (supporting / cover) was calculated for each variation. The distribution of the relative number of reads supporting each variant was compared to a distribution derived from a cohort of control samples. Contamination was detected as an increase in variations with a relative number of supporting reads below 35%.

The system we have developed is a systematic approach for identifying contamination in samples used in next-generation sequencing experiments. Contamination presents as a substantial and distinctive increase in the fraction of variations for which the relative fraction of supporting reads is below 50%. We have evaluated several dozen of our own exomes, as well as thousands of publicly available exomes, for evidence of sample-level contamination. Our analysis correctly identified two positive controls (known/ intentional contaminated samples) within our own dataset. In addition, we identified several exomes from publicly available datasets that show all of the hallmarks of sample contamination. Together with our collaborators we will be validating samples that appear contaminated to evaluate our algorithm’s specificity and sensitivity.

We have developed a simple method for identifying contaminated samples in exome sequencing experiments. Further research in this area is needed to determine the power of this method in identifying and quantifying the extent of contamination, and the amount of contamination that can be tolerated without compromising accuracy.

1556F New Mixed Model Estimates Drug-Effect or Disease Onset in Association with Covariates. M. Xu, Y.Y. Shugart. Unit on Statistical Genomics, Intramural Research Program, National Mental Health, Bethesda, MD.

Many statistical models have been developed to associate human diseases with genetic or environmental factors. Previously, the disease is often described by only one measurement. Until recently, studies seeking to associate genetic and environmental factors with human disease have relied on disease phenotypes characterized by only one measurement. The task of determining how longitudinal measurements are associated with a disease’s driving factor and with other covariates requires more sophisticated type of analysis. To meet these needs, we have developed a ‘mixed model’ approach, that not only pinpoints the turning-point when disease becomes manifest or therapeutic drugs begin to take effect but also delineates associations over time with such potentially confounding factors as genotype and disease subtype. Our new model therefore describes covariance for repeated measurements over time and incorporates a correlation structure for a stationary time series. The phenotypic data used here include sequenced measurements taken during the period before a drug effect or disease symptom becomes manifest, and also after the first effects or symptoms appear. It should be noted that for most diseases and drugs, the precise timing of onset marking the change between these two sequences is not known. Inasmuch as sequenced data covering the whole time course from naivé to full-blown effects is more readily available for drug effects than for the development of diseases, we demonstrated our new model by estimating the distribution of drug effect onset and analyzing these estimated time points’ association with key covariate factors. Using a pilot data set, we conducted data-driven computational experiments and evaluative simulations to test our new framework’s ability to estimate the onset of a drug’s effects. The working example we will present demonstrates that our new statistical framework successfully predicted the onset of a drug’s effects on baseline anxiety in a subgroup of STAR*D participants. Our results also confirm that the new approach can calculate the required depth for a targeted power with only a nominal rate of type I errors. Finally, our demonstration shows that our new statistic, researchers including those working with publicly available data--to pinpoint the onset of a specific disease or a specific drug and to determine how onset may be affected by such potentially influential covariates such as genotypes or sub-phenotypes.

1557W Fast and Accurate Diploid Genotype Imputation via Segmental Hidden Markov Model. L. Zhang1, Y.F. Pei1, H.W. Deng1, Y.P. Wang1. 1) University of Shanghai for Science and Technology, Shanghai, Shanghai, China; 2) Biostatistics and bioinformatics, Tulane university, New orleans, LA.

Fast and accurate genotype imputation is necessary for facilitating gene mapping studies, especially with the ever increasing numbers of both common and rare variants generated by high throughput sequencing experiments. However, most of the existing imputation approaches suffer from either inaccurate results or heavy computational demand. In this study, aiming to perform fast and accurate imputation analysis, we propose a novel, fast and yet accurate method to impute diploid genotypes. Specifically, we extend a hidden Markov model that is widely used to describe haplotype structures. But we model hidden states onto single reference haplotypes rather than onto pairs of haplotypes. Consequently the computational complexity is linear to size of reference haplotypes. We further develop an algorithm to speed up the calculation. Working on compact representation of segmental reference haplotypes, the algorithm always calculates an exact form of transition probabilities regardless of partition of segments. Both simulation studies and real data analyses demonstrated that our proposed method was comparable to most of the existing popular methods in terms of imputation accuracy, but was much more efficient in terms of computation. The developed algorithm can further speed up the calculation by several folds without loss of accuracy. The proposed method will be useful in large scale imputation studies with a large number of reference subjects. The implemented multi-threading software FISH is publicly available.

1558F CeRNA interactions cooperate with genomic variability to modulate drivers of tumorigenesis. H.S. Chiu1,2, X. Yang1,2,3, M.R. Martinez1,2,3, P. Sumazin1,2,3, A. Califano1,2,3. 1) Columbia Initiative in Systems Biology, Columbia University, 630 West 168th Street, New York, NY 10032; 2) Center for Computational Biology and Bioinformatics, Columbia University, 630 West 168th Street, New York, NY 10032; 3) Department of Biomedical Informatics, Columbia University, 630 West 168th Street, New York, NY 10032.

Recent evidence suggests that RNAs compete for binding and regulation by a finite pool of microRNAs (miRNAs), thus regulating each other through a competitive endogenous RNA (ceRNA) mechanism. Using a kinetic model, we show that pathophysiologically relevant regulation occurs when ceRNA interactions are mediated by multiple miRNAs, while interactions mediated by a single miRNA may have negligible magnitude. Furthermore, our model predicts that ceRNA interactions mediated by multiple miRNAs are largely independent of the individual miRNA’s expression and thus highly conserved across distinct cellular states. Indeed, we predict and validate an ultra-conserved network that includes more than 160,000 ceRNA interactions, which are conserved across tumor and non-tumor related cellular contexts. We show that this network integrates genetic and epigenetic alterations of cognate ceRNA regulators to dysregulate established oncogenes and tumor suppressors, accounting for a large fraction of the missing genomic variability in tumors.
1559F MITOCHONDRIAL DISEASE SEQUENCE DATA RESOURCE (MSeqDR) CONSORTIUM: A Global Grass-Roots Effort to Compile, Organize, Annotate, and Analyze Whole Exome and/or Genome Datasets from Individuals with Suspected Mitochondrial Disease. X. Gai1, D. Krotoski2, M. Gonzalez3, D.C. Wallace4, M. Parisi5, S. Zuchner2, M.J. Falk2. 1) Center for Biomedical Informatics, Loyola University Stritch School of Medicine, Maywood, IL 60153; 2) NICHD, NIH, Bethesda, MD 20892; 3) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL 33136; 4) Center for Mitochondrial Medicine, The Children’s Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine, Philadelphia, PA. 1598F Division of Cardiology, Beth Israel Deaconess Medical Center, Boston, MA 02215. MSeqDR prototype development is now under way with support from the UMDF and North American Mitochondrial Disease Consortium involving investigators from 9 countries in North America, Europe, Australia, and Asia. A global MSeqDR will fill the existing void in bioinformatics tools and knowledge necessary for efficient WES data analysis in the mitochondrial disease community. This resource will facilitate support reliable genetic diagnosis in individual cases, facilitate identification of additional cases of rare genetic disorders, provide evidence to exclude from consideration sequence variants of unknown significance that are unlikely to be disease-causing, permit analysis of modifier genetic factors underlying disease variability, and even identify genetic variants that are likely to alter response to emerging therapies.

1561T Comparison of the performance of read mappers and assembly methods on indel calling. H. Lin, Y. Shen. Columbia University, New York, NY. The primary technique for detecting genomic variations from next generation sequencing data is to map reads to a reference genome and call potential variants based on the aligned reads. Although this approach has worked well for detecting SNP’s, the problem of detecting insertions and deletions from short read data remains challenging, especially if the insertions and deletions are large. A number of new techniques have been developed for mapping reads more effectively in the presence of insertions and deletions. In this work, we compare the performance of a number of new read mappers and their ability to detect indels (i.e. insertions and deletions) from real sequencing data. Our data consists of sequencing from trios consisting of a proband and his or her parents. This trio data provides a unique opportunity to evaluate the quality of the indels called from different read mappers, in addition to measuring the overall number of indels called. In general, measuring the accuracy of indels called from real sequencing data can be a challenge, but with sequencing from trios, we can measure the mendelian concordance and the transmission ratio of the indels called, which provides a reasonable measure of precision. In addition, for each read mapper, we also measure the number of novel indels called, which are not found in existing databases, such as dbSNP, and also measure their mendelian concordance and transmission ratios. Lastly, we also explore the performance of some newly developed assembly based approaches for indel calling, and the impact of error correction on the methods.

1562F A Bayesian framework for de novo mutations calling in trios. Q. Wei1, Y. Liu2, Y. Han3, X. Zhan2, W. Chen2, B. Li3. 1) Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 2) Center for Human Genetic Variation, Duke University, Durham, NC; 3) Department of Biostatistics, University of Michigan, MI. For most human complex diseases, although numerous loci have been identified, the genetic basis remains largely unknown. For sporadic cases without family history, increasing evidence shows that de novo mutations play an important role in the genetics of these diseases such as neurodevelopmental disorders. Identifying de novo mutations in such cases will not only help directly find genes implicated in disease etiology but also provided candidate genes or pathways for mapping/prioritizing variants associated with the disease. High-throughput next-generation sequencing enables a genome- or exome-wide detection of de novo mutations by sequencing proband-parents trios. Since de novo mutations are extremely rare, e.g. 50-100 per generation, it is challenging to sifttrue mutations through sequencing error and alignment artifacts. Traditional approaches infer individual genotypes, and identify putative de novo mutations by comparing proband and parental genotypes. Recently an efficient likelihood-based framework (polymutt) was proposed and shown to outperform standard approaches dramatically. A limitation of polymutt is that the pre-specified prior mutation rate has a significant impact on the de novo mutation calling, resulting irreduction of sensitivity and specificity when improper priors are assigned. To address this problem, we developed triodenovo, a Bayesian framework that separates de novo calling from the mutation rate for trio data so that the prior mutation rate can be adjust post-hoc. Through extensive simulations we showed that this new method has higher sensitivity and specificity than polymutt and standard approaches, especially when the depth of coverage is low (e.g. 20x). Coupled with machine learning approaches to filtering false positive candidates due to alignment artifacts, we showed that triodenovoachieves increased specificity without sacrificing sensitivity on real data. We hope that this new framework is useful to the research community to efficiently identify de novo mutations to facilitate the association mapping of genes for human diseases.
1563W  
PhenoExplorer: a tool to help researchers identify relevant studies and phenotypic variables in dbGaP. J. Ambite1, L. Lange2, S. Sharma3, S. Voinea1, C. Hieu4, Y. Anh6, 1) Information Sciences Institute, University of Southern California, Marina del Rey, CA; 2) Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC.

The National Institutes of Health (NIH) created the dbGaP resource as a repository for genetic and phenotypic data. dbGaP provides the scientific community an unprecedented opportunity to gain access to data from multiple studies, but identifying the pertinent subset of studies and phenotype variables in dbGaP is often challenging. The true utility of this resource is thus limited by the inability to identify relevant studies on your own research. Some studies include thousands of phenotype variables, many being study-derived variables measuring similar traits, with descriptions varying widely across studies. For example, hypertension status is described variously as "HTN", "high blood pressure", and "high BP", causing many exact text searches to fail to identify all relevant variables. Moreover, a search for hormone replacement therapy should also return studies with variables about estrogen intake, which requires a semantic search. In addition, users often wish to limit phenotype searches to studies with certain features, such as those with African Americans and/or with sequence databases. To address these issues, we developed PhenoExplorer.org, a free online tool that allows researchers to identify dbGaP studies containing phenotypes of interest and the corresponding phenotypic variables. Specifically, a researcher can search for studies along a set of dimensions, including race/ethnicity, sex, study design, type of genetic data, genotype platform, and diseases studied. Crucially, the researcher can also enter free text searching the phenotype of interest. The tool returns all studies satisfying the metadata constraints and matching the free text phenotype search. PhenoExplorer.org uses an ensemble of semantic similarity and text similarity techniques (TF-IDF, cosine similarity). In summary, PhenoExplorer enhances access value to our research efforts.

1565F  
Reconstructing Pedigrees from Estimates of Genomic Sharing in Admixed Populations. J.E. Below1, J. Staples2, A. Reiner3, L. Ekybekova4,5, S. Musani2, J.G. Wilson2, C.R. Lin4, D. Nickerson1. 1) Human Genetics, University of Texas Health Science Center, Houston, TX; 2) University of Washington, Genome Sciences, Seattle, WA; 3) University of Washington, School of Public Health, Seattle, WA; 4) Jackson State University, Jackson, MS; 5) University of Mississippi Medical Center, Jackson, MS.

Understanding and correctly utilizing relatedness among samples is essential for all genetic analysis. However, records of sample relatedness are often missing or unavailable. PRIMUS is an algorithm that utilizes genome-wide estimates of identity by descent (IBD) to assign relationship categories and leverages these pairwise relationships to identify all possible pedigrees consistent with the observed genetic sharing. Reconstructed pedigrees in admixed populations is complicated by the fact that many algorithms for estimating IBD from genome-wide SNP data assume sampling from a single homogeneous ancestral population. The presence of ancestry informative markers (AIMs) in estimating IBD using a method of moments can strongly bias relationship classification, resulting in inaccurate or failed pedigree reconstruction. We have implemented a principal component analysis within PRIMUS to identify signatures of admixture, appropriate reference population minor allele frequencies, and identity and remove AIMs prior to IBD estimation. Controlling for multiple ancestral groups, we reconstructed pedigrees for all 1,985 Mexican Americans from the Starr County Health Study (SC) as well as 3030 African Americans from the Jackson Heart Study (JHS). PRIMUS unambiguously identified 197 previously undescended pedigrees in SC and reconstructed pedigrees for 158 pedigrees in JHS. We conclude that PRIMUS can successfully reconstruct pedigrees in genetically heterogeneous samples. We present data for the resulting pedigrees, and show that PRIMUS is powerful for both identifying novel pedigrees in large admixed genetic cohorts and for validating known pedigrees.

1566W  
Detecting Differentially Expressed Genes in RNA-Seq Data with Unknown Conditions. G. Klambauer1, T. Unterrhiner2, S. Hochreiter. 1) Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Upper Austria, Austria;

Methods that identify differential expression in RNA-Seq data are currently limited to study designs in which two or more sample conditions are known a priori. However, these biological conditions like activated regulatory and metabolic pathways are typically unknown in genetic studies such as the HapMap or the 100 Genomes project. We suggest DEXUS for detecting differential expression in RNA-Seq data for which the sample conditions are unknown. In a Bayesian framework DEXUS models read counts as a finite mixture of negative binomial distributions in which each mixture component corresponds to a condition. Evidence of differential expression is measured by the informative/non-informative (I/NI) value, which allows differentially expressed transcipts to be extracted at a desired specificity (significance threshold) or a desired sensitivity (false discovery rate). In cohorts with genetic and RNA-Seq data, DEXUS was able to detect differentially expressed transcripts that could be related to genetic variants via the identified conditions. These genetic variants can be classified into structural variants like copy number variations and single nucleotide variants, that is, eQTLs.

1567T  

The use of next-generation sequencing (NGS) as a tool for interrogation of somatic/germline variation in DNA and differential expression in RNA pervades many areas of biology. With the adoption of this technology has come the ability to detect many different types of experiential variation that were previously not possible. However, the advent of high-throughput NGS processes for RNASeq is very different than that for DNASeq, which leads to different challenges. We have developed rapid and sensitive methods to detect and quantitate the sources of variation which occur during the ‘wet-lab’ stage of NGS analysis. Our methods include methods to detect and remove contaminating sequences, detect 3’ bias often observed in RNA sequencing data, and resolve sample identity issues. These processes are providing critical information about the quality of samples generated from internal projects (as well data acquired from external sources) which is providing significant value to our research efforts.
RNA-sequencing workflow: tool integration and visualization of GWAS data.

RNA-sequencing workflow to fully leverage multiple processors. An extensive workflow that takes the utmost advantage from an RNA-seq experiment and provides comprehensive reports on genomic features for secondary data analysis. MAPRseq workflow integrates a suite of open-source bioinformatics tools along with in-house developed methods to analyze paired-end RNA-seq data. Read alignment is performed with Tophat which uses Bowtie - a fast, memory efficient, short sequence aligner. Tophat aligns reads to the transcriptome and further to the genome to report both existing and novel junctions. Along with the alignment (BAM) and junction (BED) files, Tophat also provides a list of expressed fusion transcripts using the TopHat-Fusion algorithm. The BAM file is processed using HTSeq to summarize expression at gene level. Exon quantification is obtained with in-house methods that leverage BEDTools. In addition to raw gene and exon expression counts, MAPRseq also provides normalized values (RPKM). For accurate variant detection, GATK is used to call SNVs that are further annotated with qualitative results, coverage and additional criteria using VCFR. MAPRseq workflow reports several analytical functions, including alignment statistics, in-depth quality control metrics, gene and exon expression levels, fusion transcripts and SNVs for each sample. Circos plots are also provided to visualize fusion transcripts. MAPRseq incorporates Integrated Genomics Viewer (IGV) to visualize alignment and coverage along the transcriptome as well as exon-exon junctions. MAPRseq incorporates UCSC tracks on transcription and regulation in IGV as well to facilitate user interpretation. Library details (library design, library preparation, library sequencing, etc.) can be examined in MAPRseq. The workflow is fully optimized to run on a single Linux machine as well as in a cluster to enable it to support a single user to leverage multiple processors.

Assessment of the impact of read length on RNA-seq results: An ABRF Consortium study. The ABRF-NGS Consortium aims to evaluate the performance of NGS platforms and to identify optimal methods and best practices. As part of this evaluation, we analyzed the utility of short vs. long reads and single vs. paired reads from each end of a template, for identifying differentially expressed genes and novel isoforms. Read length and pairing are critical parameters affecting experimental design, analysis, and cost, and longest available reads are often used in paired-end mode by default, whereas most experimental choices are necessary for the accurate RNA-seq profiling. To test this question, we used two standard ABRF RNA-seq libraries that were used throughout the ABRF-NGS project and the preceding MAQC project. To confirm our results on an independent dataset, we utilized data from ENCODE on 2 cell lines. In order to mitigate the influence of a particular computational algorithm, we utilized 3 separate pipelines to analyze the data. In order to further control the data, we used paired-end 100bp reads and computationally produced synthetic shorter and single and reads from these real reads by trimming their length. The results were extremely consistent, but not in a positive way. We found that read-length and paired or single-end status are important factors in determining the results of an RNA-seq experiment. We found that the number of differentially expressed (DE) genes varied greatly as the read length changed. For many of the genes, there was a slight shift in FPKM, which was sufficient to move a hit from being DE to being insignificant. This study indicates that while RNA-seq is a powerful method which can produce useful results, the current technique yields a substantial variability in results that cannot be dismissed. Critical improvements are needed in RNA-seq sample preparation, data generation, and bioinformatics analysis tools.

Enlight: a web-based tool for integrating GWAS results with biological annotations.

Enlight is a powerful method which can produce useful results, the current technique yields a substantial variability in results that cannot be dismissed. Critical improvements are needed in RNA-seq sample preparation, data generation, and bioinformatics analysis tools.

Enlight: a web-based tool for integrating GWAS results with biological annotations.

Enlight is a powerful method which can produce useful results, the current technique yields a substantial variability in results that cannot be dismissed. Critical improvements are needed in RNA-seq sample preparation, data generation, and bioinformatics analysis tools.
Network Communicability: An Effective Alternative Metric for Genome Analysis. C. Shaw, I.M. Campbell, Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX.

Network similarity metrics are increasingly utilized to productively analyze genome-wide data. Conventional approaches such as shortest path and clique-based techniques have been useful but are not well suited to all applications. Recently, computational scientists in other disciplines have developed network communicability as a complementary method to analyze network relationships. Network communicability considers all paths between two network members. Given the success of previous analyses of the human protein-protein interaction network, we sought to apply network communicability to this complex data. As a challenging test case, we sought to partition human protein products into disease classes using communicability together with a set of training genes with known disease annotations based on OMIM. Our analysis reveals that communicability has several key advantages over shortest path and other previous methods. Additionally, we used copy number variation data from our diagnostic lab to show that the disease classes of genes deleted in patients correlate with their respective phenotypes. Our data suggest that metrics based on network communicability have considerable utility in the analysis of large-scale biological networks and may be fruitfully applied to a number of computational problems.

The genome of a species typically encompasses great diversity, reflecting the mutation processes and selective forces experienced within its evolutionary history. A single sample may be used as a canonical exemplar, leading to a reference genome as the basis for analysis. This is a good first approximation and is the basis of sequence assembly for most high throughput sequencing experiments. However, this approach typically fails in regions where the sample is highly diverged from the reference and does not take advantage of known variants. We have developed an approach that combines multiple reference sequences, previously known variations and de novo assembly for inferring genome sequence. The first step is to infer a diploid reference using a Hidden Markov Model on an ordered graph representing a multiple sequence alignment of known references. These may be entire genomes, long haplotypes, divergent exons or short variants and are augmented with sequence from the sample under study. A diploid genome corresponds to a paired path through the graph. The second step is to map reads to the pair of personal references and infer variants by standard approaches. By doing so, we improve sensitivity to novel variant detection, particularly in regions that diverge from the standard single reference approach. We demonstrate the value of the personal reference approach in the analysis of the human Major Histocompatibility Complex, particularly in its ability to characterise structural variation and regions of extreme polymorphism frequently found in this region.
1579T Multi-platform and cross-methodological reproducibility of transcriptome profiling by RNA-seq in the ABRF Next-Generation Sequencing Study (ABRF-NGS). C.E. Mason1,2, E.M. Cenzer3, C.W. Tignon3, G. Roberson3, D. Grove1, S. Levy4, W. Farmerie5, A. Pale1, C. Wright6, P.A. Schweitzer10, Y. Gao10, D. Kim11, B. Holz12, B. Hicks13, R. Kim13, S. Chiang14, N. Jafari14, N. Raghuvar15, C. Hendrickson15, D. Roberson15, J. Rosenfeld15, T. Smith15, J. Underwood15, M. Wang15, P. Zumbo1, D. Baldwin15, G. Grills10, ABRF-NGS Consortium. 1) Department of Pathology and Anatomy, Weill Cornell Medical College, New York, New York, USA; 2) The HHPrince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biology, Weill Cornell Medical College, New York, New York, USA; 3) Vermont Cancer Center, University of Vermont, Burlington, Vermont, USA; 4) Keck Medical Center, University of Southern California, Los Angeles, California, USA; 5) The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, USA; 6) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA; 7) Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, Florida, USA; 8) Memorial Sloan-Kettering Cancer Institute, New York, New York, USA; 9) Roy J. Carver Biotechnology Center, University of Illinois, Urbana, Illinois, USA; 10) Biotechnology Resource Center, Institute of Biotechnology, Cornell University, Ithaca, New York, USA; 11) Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA; 12) NIH/NCI/SAIC-Frederick, Gaithersburg, Maryland, USA; 13) Genome Center, University of California, Davis, Davis, California, USA; 14) Center for Genetic Medicine, Northwestern University, Chicago, Illinois, USA; 15) NIH/NHLBI, Bethesda, Maryland, USA; 16) Division of High Performance and Research Computing, University of Medicine and Dentistry of New Jersey, Newark, New Jersey, USA; 17) PerkinElmer Inc., Seattle, Washington, USA; 18) Pacific Biosciences, Menlo Park, California, USA; 19) Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, USA; 20) Pathomarics LLC, Philadelphia, Pennsylvania, USA.

Next-generation sequencing (NGS) technology applications such as RNA-seq and whole-genome sequencing (WGS) have been extensively used in clinical diagnostics and research for a variety of applications including disease diagnosis and personalized medicine. However, reproducibility and accuracy of NGS data are critical for reliable interpretation of results. In this study, we compared the reproducibility and accuracy of RNA-seq data across multiple platforms and methods, including Illumina HiSeq 2500/2500, Life Technologies PGM/Proton, Roche 454 GS FLX+, Complete Genomics (CG) platforms, sequenced at average coverage of 26x and 40x, and processed with BWA plus GATK and CG proprietary pipeline, respectively. The effect of various filters on the platform bias was evaluated. As expected, the bias was much larger concerning indels (at least 20% of the variants after initial QC) than compared to single-nucleotide polymorphism (SNP) calls, however, significant discrepancy was detected also for the latter (at least 5% of the variants after initial QC). It is known that the next-generation sequencing errors are mostly due to mismsapping of the sequence reads. Recently Lee and Schatz (Bioinformatics 28, 2097-2105) and Derrien and others (PLOS One 7: e30377) proposed tools that can help to avoid such errors by identifying and excluding regions of the genome with low mappability. We built on similar in-house efforts. Applying the resulting poor-mapping windows together with adjusted quality and coverage filters, we were able to minimize the platform bias significantly. We work toward evaluating the effect of other sequence read mappers (Bowtie2, GATK2, etc.).

1581W Reducing platform bias in next-generation sequencing. L. Saag1,2, U. Gerst Talas1, M. Mitt1, R. Villerm1, M. Metspalu1,2. 1) University of Tartu, Tartu, Estonia; 2) Estonian Biocentre, Tartu, Estonia.

Next-generation sequencing platforms enable re-sequencing of the human genome at high speed and relatively low cost. At the same time, there are discrepancies in base calls from different sources - the platform bias. The relatively low concordance between the results from different sequencing platforms has been reported (Lam et al. Nature Biotechnology 30, 78-82; Ratan et al. PLOS One 8: e55089). This can be a serious problem when data from various sources need to be pooled, especially, for example, in cases where different populations (or cases and controls) are sequenced with different technologies. We compared the concordance of the base calls from four human genomes, produced on the widely utilized Illumina and Complete Genomics (CG) platforms, sequenced at average coverage of 26x and 40x, and processed with BWA plus GATK and CG proprietary pipeline, respectively. The effect of various filters on the platform bias was evaluated. As expected, the bias was much larger concerning indels (at least 20% of the variants after initial QC) than compared to single nucleotide polymorphism (SNP) calls, however, significant discrepancy was detected also for the latter (at least 5% of the variants after initial QC). It is known that the next-generation sequencing errors are mostly due to mismsapping of the sequence reads. Recently Lee and Schatz (Bioinformatics 28, 2097-2105) and Derrien and others (PLOS One 7: e30377) proposed tools that can help to avoid such errors by identifying and excluding regions of the genome with low mappability. We built on similar in-house efforts. Applying the resulting poor-mapping windows together with adjusted quality and coverage filters, we were able to minimize the platform bias significantly. We work toward evaluating the effect of other sequence read mappers (Bowtie2, GATK2, etc.).

1582T Customized and Personalized Next Generation Genomics. A.N. Singh. Computer Science, Virginia Tech, Blacksburg, VA.

Structural variations, SVs, with size 1 base-pair to several 1000s of base-pairs with their precise breakpoints and single-nucleotide polymorphisms, SNPs, were determined for members of a family of four. It is also discovered that the mitochondrial DNA is less prone to SVs re-arrangements than mtDNA. In addition, the number and proportion of SVs was determined. We have developed a customized alignment approach that uses the re-arranged segments. This can help to avoid such errors by identifying and excluding regions of the genome with low mappability. We built on similar in-house efforts. Applying the resulting poor-mapping windows together with adjusted quality and coverage filters, we were able to minimize the platform bias significantly. We work toward evaluating the effect of other sequence read mappers (Bowtie2, GATK2, etc.).

1588OF Identifying Genomic Copy Number Alteration and Loss of Heterozygosity in Next-Generation Sequence Data. S. Rozen1,2, J.R. McPherson1, Y. Wu1,4, P. Tan1,2,3,4. 1) Duke-NUS Graduate Med Sch, Singapore, Singapore; 2) Singapore-MIT Alliance in Research and Technology; 3) National Cancer Centre, Singapore; 4) Genome Institute of Singapore.

Losses and duplications of large genomic regions resulting in copy number alterations or loss of heterozygosity (LOH) are common drivers of cancer development. Until recently, these aberrations have been characterized primarily by means of SNP and copy-number oligonucleotide arrays. However, next-generation sequencing has become a dominant approach for detecting mutations in cancer genomes, and the read data generated in these studies can also be used for characterizing genomic aberrations in cancers. We developed an analytical approach and software that considers read depth and read counts for each allele to identify genomic regions with copy number alteration or LOH. The approach, RDAAC (Read Depth and Allele Counts) requires aligned sequencing reads from matched malignant and non-malignant DNA. RDAAC simultaneously estimates the proportion of non-tumor genomes in the sample and the number of copies of each SNP allele in the tumor sample. We assessed RDAAC’s performance in two ways. First, we compared its results to results from analyzing SNP arrays. Second, we assessed its performance on simulated data, which are an example of a genome with copy number alternations, regions of LOH, and admixture of non-malignant DNA were known. RDAAC functioned well on whole-exome sequencing data even in the presence of generalized polyploidy and >50% admixture of non-malignant cells. We developed R package that implements (1) RDAAC on top of SAMTools(Li et al. 2009) and ASCAT (which was designed for microarray data, Van Loo et al., 2010) and (2) the simulator for synthetic data.
Identifying Mendelian Disease Genes; an Analysis Tool of PhenoDB. N. Sobreira1, F. Schiettecatte2, D. Valle3, A. Hamosh1. 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) FS Consulting, Salem, MA.

WES has been the main method used to search for Mendelian disease genes in the last 3 years. But identifying the pathogenic mutation among thousands to millions of genomic variants is a challenge, and variant prioritization strategies are required. The choice of these strategies depends on the availability of well-phenotyped patients and family members, the mode of inheritance, the severity of the disease and its population frequency. Here we describe the PhenoDB analysis tool developed to rapidly identify a list of candidate variants by applying various strategies for disease variant prioritization. PhenoDB is a Web-based portal for managing and analyzing phenotypic/clinical information and WES/WGS data. The VCF and ANNOVAR files, genotyping array data, QC report, and final results file are available in the Analysis module, in which the user can choose the family members to include, the inheritance mode, what kind of variants to filter on (missense, nonsense, splicing, 3’UTR, 5’UTR, synonymous), optionally exclude variants found in the dbSNP 126, 129 and/or 131, the MAF value for exclusion of variants in the 1000 genome and EVS databases, and exclusion of chrX variants. The analysis runs in less than 1 minute and generates a list of candidate genes plus a log of the prioritization strategy. We also add information from external databases such as OMIM, MGI, GeneCards, etc. Multiple analyses with different prioritization strategies can be saved. Any saved ANNOVAR file or analysis result can be filtered using the 57 genes listed in the ACMG incidental findings list or compared against the genes in an OMIM phenotypic series. The user can also compare analysis results across families for overlapping mutated genes and cohorts (with or without locus heterogeneity) can be analyzed. A Sandbox is provided where ANNOVAR files from families not in PhenoDB can be uploaded and analyzed. Genes and variants can be searched across families in the analysis results. The PhenoDB analysis tool has been proved to be easy, fast and efficient. In the Baylor-Hopkins Center for Mendelian Genomics we have now analyzed 83 families using this tool. We have identified the causative gene in 14 families and suggested candidate genes that are being further investigated in 39 other families. PhenoDB itself is freely available through the Johns Hopkins McKusick-Nathans Institute of Genetic Medicine (http://phenodb.net).

The European Genome-phenome Archive (EGA). V. Kumanduri, A. Datta, J. Almeida-King, L. Clarke, I. Lappalainen, P. Flicek, J. Paschall, EMBL-EBI, EMBL-EBI European Bioinformatics Institute, Hinxton, United Kingdom.

The European Genome-phenome Archive (EGA) provides a permanent archive for genetic and phenotypic data and a secure means to share subject level data. EGA contains data collected from individuals for the purpose of medical or genetic research and whose consent agreements require security and request management. Data access decisions are not made by the EGA but by a submitter defined Data Access Committee (DAC). Once DAC authorization to the data has been granted, the EGA provides secure tools for downloading and decrypting the authorized files and general support for questions related to the data content. The EGA is available at http://www.ebi.ac.uk/ega/. Data from more than 350 studies, comprising 500 datasets and 200,000 individuals have been submitted to the EGA and is made available to authorized researchers. These studies include data from array-based genotyping experiments as well as raw DNA sequence data from re-sequencing or transcriptomics projects. While continuing to perform the main function of archiving datasets exactly as submitted, EGA is now developing an option for submitters to choose value-added processing steps. Specifically, we aim to apply a defined analysis pipeline to provide standardized variant call sets that can aid in comparisons of methods and in the combining of data from multiple sources. In order to achieve this, the EGA is building a data analysis workflow that allows to align reads against a reference, provide high quality variant calling and distribute phased genotype information for the analysed samples. For those submitters who request this standardized processing, these features will be run in addition to archiving of the primary data files and analysis products. Anyone interested in further information about the EGA in general should contact ega-helpdesk@ebi.ac.uk.
Scalable variant identification and imputation across large multigenerational pedigrees from high-throughput sequencing data by joint Bayesian variance calling. F.M. De La Vega 1, S. Malakalsh 2, R. Litinin 3, L. Triggs 4, A. Jackson 5, D. Ware 6, J.G. Cleary 7.

Whole-genome sequencing (WGS) has revolutionized biological studies to identify variants for rare diseases and complex traits. Often these studies sequence related individuals to varying sequencing depths across a pedigree. However, variant calling methods for high-throughput sequencing (HTS) pipelines or thRN (htSNP) array methods lack scalability to pedigrees. We present a Bayesian framework that jointly calls variants across a pedigree implicitly leveraging shared haplotypes. When calling variants at a given position for a nuclear family, we look-up the alignment data of all family members simultaneously, scoring genotypes across pedigree members with priors based on Mendel's laws of variant segregation, and handling sex chromosomes as special cases. As validation, we analyzed data from a 17 individual, 3-generation CEPH pedigree sequenced to 40X average depth. We focus our analysis on NA12878, a female in the second generation and for which extensive orthogonal validation data is available from the 1000 Genomes Project and other efforts. Compared to singleton calling, our family caller produced more high quality variants and eliminated spurious calls judged by standard quality metrics such as T/T Het/Hom ratios, and dbSNP/OMNI array data concordance. Through the analysis of the segregation of variants to the 11 offspring of the third generation called independently, we estimate less than 1%, 2.5% and 2.75% of false positive calls for SNVs, indels, and MNP, respectively. Family designs should allow identification of de novo mutations, which are of great interest in neurodevelopmental and other disorders. By allowing a small prior for de novo mutations and scoring specifically for this type of mutations, we observed 100% and 98% sensitivity in identifying previously validated germline and somatic de novo mutations, respectively, compared to a de novo detector which is a rate of reduction far greater than that observed in Mendelian disorders. Simple Mendelian scoring would scale exponentially across multi-generational pedigrees. Instead, we implemented a method for acyclic prior propagation beyond nuclear families, approximating the full pedigrees in a tractable fashion in linear time. Our results show that our joint pedigree calling method outperforms singleton and population variant calling in pedigrees, allows for the identification of de novo mutations with greater specificity, and is scalable to large genomes and human disease studies.
1590W


The utility of long sequencing reads for enhanced genomic annotation and assembly has been a key differentiator for successful next generation sequencing applications. Through optimization of both the sequencing and amplification biochemistries, we have significantly improved Ion Torrent’s semiconductor chip-based sequencing system. Along with improvements in engineering and software, these innovations result in robust high quality 400-base sequencing reads on the Ion Torrent PGM™ platform. Using this improved 400 bp biochemistry on a 318 chip, we have produced >3G aligned Q20 bases with even coverage on the Escherichia coli genome. Additionally, with optimized emPCR conditions we have rapidly improved coverage uniformity of the more challenging Rhodobacter sphaeroides genome from 88% to 95%. Compared to the previous PGM™ sequencing protocols, the Ion Torrent 400bp Template/Sequencing kits give nearly double the read length and throughput. Such improvements in the PGM® system enable a broader range of applications, such as enhanced de novo genome assemblies, Human Leukocyte Antigen (HLA) sequencing, bacterial identification, and meta-genomic analysis. For Research Use Only. Not for use in diagnostic procedures.

1591T


High-throughput sequencing (HTS) has recently provided price competitive alternatives to microarrays for both RNA expression profiling with the RNA-seq protocol and DNA genotyping with whole genome and whole exome sequencing. Although the bioinformatics tools have matured for secondary analysis of sequence data, including alignment, variant calling, and gene and transcript level quantification, the outputs of these tools often require inspecting the ‘raw read alignments’ for putative variants and genes with interesting expression profiles. Investigating these variants in their VCF format and the alignments in BAM format allows for detection of false-positives as well as aiding the interpretation process by providing a rich genomic context.

We introduce Golden Helix GenomeBrowse™, a free visualization tool for DNA and RNA sequence alignment and variant calls along with annotations tracks from a rich catalog of pre-curated public data. GenomeBrowse is built from the ground up with the guiding principles of (1) working seamlessly with cloud-hosted data as fluidly and quickly as local files, (2) being intuitive to use for non-bioinformaticians to utilize in their research, and (3) having a multi-threaded architecture to make big-data visualization and analysis accessible to anybody capable of navigating Google Earth. By including integration with a rich repository of public data, users have no barriers to the process of interpreting their sequencing results. In particular, we demonstrate the ability of GenomeBrowse to stream exome sequencing of a trio from the Amazon Cloud from whole genome views down to the gene level with annotation tracks ranging from 1000 Genomes, dbSNP, genes, and miRNAs. We show how GenomeBrowse can highlight false-positive Single Nucleotide Variants and small Insertion/Deletions, confirm the inheritance pattern of putative functional variants, and aid in the interpretation of a variant’s impact.

1592F

Improving the robustness of personal genome variant discovery: the impact of technical replicates, sample source and analysis method. D. Mittelman1, A. Del Dicera2, R.M. Iwasow2, N. Leibovich1, J. Wang2, M. Tayeb2. 1) Golden Helix, Bozeman, MT. 2) DNA Genotek Inc. A wholly owned subsidiary of OraSure Technologies Inc., Ottawa, ON, Canada K2K 1L1; 3) Arpeggi, Inc., Austin, TX, USA 78749.

Background: Whole genome sequencing (WGS) presents unique challenges to disrupt prognostic and diagnostic genetic testing in the clinic. Although the adoption of WGS is accelerating, the standards and acceptance criteria for clinical WGS are not fully defined. Variant calling robustness and reproducibility of WGS analysis was assessed across routinely used biological sample types, among technical replicates, and between sequencing library preparations. Consideration was also given to sample-to-result costs. Methods: Two multigenerational families comprising a total of 7 individuals were enrolled in the study based on low (<5%), medium (6-20%) and high (>21%) levels of non-human DNA content in their saliva as measured by qPCR. Matched sample pairs of blood (EDTA, Becton Dickinson, Inc.) and saliva (OraGene, DNA Genotek Inc.) were prepared from each participant and sequenced through the Illumina Genome Network (IGN), generating a total of 20 whole genome datasets. Multiple variant discovery pipelines, including CASAVA and GATK were used to interrogate reproducibility of biological and technical replicates and Mendelian consistency in parent-offspring trios. The computational burden (i.e. cost from DNA-to-result) between sample types was also compared. Results: A direct linear correlation (R² = 0.98) between non-human DNA content in a saliva sample and the proportion of unmapped reads after read alignment was found, suggesting that reads derived from non-human DNA are often not mapped. The average difference in unmapped reads between paired blood and saliva samples from the same individual was 8.7%. On average, 4% of variant calls (SNPs and small indels) were discordant between technical replicates of blood samples. Likewise, discordant variants calls between saliva and blood averaged 4%. Finally, QC parameters, such as minimum read depth, are suggested to maximize variant calling consistency among technical replicates and across different sample types. Conclusion: Variant concordance between matched pairs of saliva and blood is equivalent to concordance between technical replicates within sample types. This strongly suggests saliva is an effective alternative to blood for whole genome sequencing analysis.

1593W


In 2012, NIST convened the Genome in a Bottle Consortium (www.genomeinabottle.org) to develop the reference materials, reference methods, and reference data needed to assess confidence in human whole genome variant calls. We will report on progress of the four consortium working groups. The Reference Material (RM) Selection and Design group has selected the first 4 genome families from the Personal Genome Project for whole genome RMs, a candidate cancer/normal cell line pair, and artificial structures for point mutations. The Measurements for Reference Material Characterization group has initiated experiments to characterize a pilot RM (NA12878) using multiple sequencing methods, other methods, and validation of selected variants using orthogonal technologies. The Bioinformatics, Data Integration, and Data Representation working group has developed data integration and analysis methods, data sharing protocols, and representation formats. The Performance Metrics and Figures of Merit group has initiated collation of proposed performance metrics and figures of merit from consortium members, leveraging existing resources. The products of these working groups will be a set of well-characterized whole genome and synthetic DNA RMs along with the methods (documentary standards) and reference data necessary for use of the RMs. These products will be designed to help enable translation of whole genome sequencing to clinical applications by providing widely accepted materials, methods, and data for performance assessment.
1594T
Change can be good: updating the human reference genome assembly.
V.A. Schneider1, P. Pilek2, T. Graves3, T. Hubbard4, D.M. Church5 for the Genome Reference Consortium. 1) NCBI/NIH, Bethesda, MD 28092; 2) EBI, Hinxton, Cambridge, CB10 1SD, UK; 3) The Genome Institute at Washington University, St. Louis, MO 63108; 4) The Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK.

Since its publication in 2003, the human reference genome assembly has become an interval reagent in biological research. In today’s era of next generation (NGS) and personal genome sequencing, it is the framework upon which individual genome analysis is based as well as a common coordinate system for genomic annotation. While the human reference retains its distinction as the highest quality mammalian genome, studies in the past decade have demonstrated that the linear chromosome model initially used for the assembly is insufficient in its ability to represent population variation. Within this model, the insertion of sequences from multiple haplotypes at complex variant locations led to non-existent allele combinations and artificial gaps. To provide a more accurate genome representation and better represent variation, the Genome Reference Consortium (GRC), the group responsible for the upkeep of the human reference assembly, developed a new assembly model. First implemented for GRCh37 (hg19), the new model retained the linear chromosomes but introduced alternate loci, scaffold sequences placed into chromosome context via alignment, as a mechanism for providing alternate sequence representation at complex or variant genomic regions. The new assembly model also introduced patches, operationally similar to the alternate loci, as a means for providing users access to assembly corrections and new variant sequence representations without coordinate-disrupting full assembly updates. Since the production of GRCh38, the GRC has released patches to 170 genomic regions, representing 3% of the chromosome sequence and adding more than 6 Mb of novel sequence and over 180 gene models not found on the chromosome assembly. This model necessitates an update to current analysis tools that cannot yet accommodate allelic duplication. We will present data demonstrating the added benefit of using the full assembly, including alternate loci and patches, with respect to gene annotation, NGS read alignment and variant analyses. Nonetheless, coordinate-changing major assembly releases and subsequent assemblies under the new model can still confound analysis. The GRC is planning to release GRCh38 in the fall of 2013. We will present our progress on the new assembly and illustrate several of its features, including the large-scale correction of erroneous bases, addition of missing sequences and retelling of misassembled regions.

1595F

RNA-Seq is a powerful technique that allows for sensitive digital quantification of transcript levels, and enables the detection of noncanonical transcription start sites as well as termination sites, alternative splice isoforms and transcript mutation and editing. Standard ‘next-generation’ RNA-sequencing approaches generally require double-stranded cDNA Synthesis, which erases RNA strand information. Strand information is important for correct annotation of novel genes, identification of antisense transcripts with potential regulatory roles, and for correct determination of gene expression levels in the presence of antisense transcripts. To obtain high quality strand specific sequence information, RNA sequencing libraries must be highly-strand-specific and highly diverse. Ideally, they should be produced at high yield and by a fast protocol. To address these challenges, we have developed the NEBNext® Ultra Directional RNA Library Prep Kit for Illumina, which is based on labeling second strand cDNA with dUTP. Using an excision to retain RNA strand information, this approach is incorporated into a streamlined workflow, enabling directional RNA library construction in ~5hrs with minimal hands on time. Here, we provide an analysis of library quality (library complexity, continuity of gene coverage, strand specificity, and 3’ and 5’ bias), as well as demonstrate compatibility with NEBNext® barcode strategy and ribosomal RNA removal.

1596W
Targeted Sequencing for Preterm Birth Associated Genes. A. Uzun1,2, I. Kurthara1, B. Mcgonigal1, J.F. Padbury1,2,3. 1) Pediatrics, Women and Infants Hospital, Providence, RI; 2) Pediatrics, Brown Medical School, Brown University, Providence, RI.

Despite significant advances in the care of pregnant women and low birth weight infants, preterm birth remains the leading cause of newborn morbidity and mortality as well as the main cause of hospitalization in the first year of life in the United States. We hypothesize that rare variants in the genes that contribute to the risk of PTB can be identified using new bioinformatics approaches coupled with high-throughput technologies applied to appropriate cohorts of patients. Understanding of the complex pattern of gene expressions, which enhance or attenuate the risk of Preterm Birth (PTB) will require new strategies. We have developed an alternative approach to identify a more manageable set of genes for preterm birth, which nonetheless incorporates some elements from the discovery in genome wide investigations. We analyzed genotype information from a large GWAS which included >660,000 SNPs from 1,000 mother and infant controls and 1,000 mother and infant preterm birth cases. We used gene set enrichment analysis (GSEA) to investigate the individual contribution(s) of biological pathways to the genetic architecture of preterm birth. We used both our curated genes and a genome-wide approach in the pathway analysis. We identified 329 genes nested within 69 pathways that are highly associated with preterm birth. For a proof of concept, we used long-range PCR to generate a DNA library from the top 11 genes from this gene set. We sequenced gDNA from 4 women with 3 generations of preterm birth and 8 term controls. High quality sequence data from well balanced pools was observed. There were an average of 22,000,000 reads from each patient, with an average of 99 perfect index reads and a Q30 of 91%. The mean Phred score for each patient was 58. A total of 2669 initial variants were reduced to 1015 following filtering for the quality indices described above. 760 were already identified in one of several database archives, including 1000 Genomes, the Exome Sequencing Project or dbSNP. 255 were unique of which 5 were exonic and predicted to be deleterious in genomic regions encompassing PTGS2 (1 deleterious), AKT1 (2 deleterious), and TP53 (1 deleterious). There have been too few samples sequenced to date for valid statistical comparisons and inferences but we will discuss the importance of not restricting the library construction to exonic sequences alone.

1597T
Well-characterized genomes for understanding genome sequencing performance: Integrating datasets from multiple sequencing platforms to form highly confident snp and indel calls. J.M. Zook1, B. Chapman2, O. Hofmann2, W. Hide2, D. Mittelman3, J. Wang4, M. Salt1, Genome in a Bottle Consortium. 1) Biosystems and Biomatics Division, National Institutes of Standards and Technology, Gaithersburg, MD; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Virginia Bioinformatics Institute, Blacksburg, VA; 4) Arpeggi, Inc., Austin, TX.

Clinical translation of genome sequencing requires methods to understand accuracy of genotype calls at millions or billions of positions across a genome. Previous work showing high discordance between sequencing methods and algorithms has highlighted the need for a highly confident set of genotypes across a whole genome that could be used as ‘truth’ for understanding accuracy. Therefore, we have developed methods to make highly confident SNP, indel, and homozygous reference genotype calls for genome. Using several methods and algorithms has highlighted the need for a highly confident set of genotypes across a whole genome. We hypothesize that rare variants in the genes that contribute to the risk of PTB can be identified using new bioinformatics approaches coupled with high-throughput technologies applied to appropriate cohorts of patients. Understanding of the complex pattern of gene expressions, which enhance or attenuate the risk of Preterm Birth (PTB) will require new strategies. We have developed an alternative approach to identify a more manageable set of genes for preterm birth, which nonetheless incorporates some elements from the discovery in genome wide investigations. We analyzed genotype information from a large GWAS which included >660,000 SNPs from 1,000 mother and infant controls and 1,000 mother and infant preterm birth cases. We used gene set enrichment analysis (GSEA) to investigate the individual contribution(s) of biological pathways to the genetic architecture of preterm birth. We used both our curated genes and a genome-wide approach in the pathway analysis. We identified 329 genes nested within 69 pathways that are highly associated with preterm birth. For a proof of concept, we used long-range PCR to generate a DNA library from the top 11 genes from this gene set. We sequenced gDNA from 4 women with 3 generations of preterm birth and 8 term controls. High quality sequence data from well balanced pools was observed. There were an average of 22,000,000 reads from each patient, with an average of 99 perfect index reads and a Q30 of 91%. The mean Phred score for each patient was 58. A total of 2669 initial variants were reduced to 1015 following filtering for the quality indices described above. 760 were already identified in one of several database archives, including 1000 Genomes, the Exome Sequencing Project or dbSNP. 255 were unique of which 5 were exonic and predicted to be deleterious in genomic regions encompassing PTGS2 (1 deleterious), AKT1 (2 deleterious), and TP53 (1 deleterious). There have been too few samples sequenced to date for valid statistical comparisons and inferences but we will discuss the importance of not restricting the library construction to exonic sequences alone.
1599W
**Power and Limitations of RNA-Sequencing**
F. Staedtler, E.J. Oakeley, Biomarker Development, Novartis Institutes for BioMedical Research (NIBR), Basel, Switzerland.

On behalf of the Sequencing Quality Control Consortium (SEQC).

In the FDA-led SEQC (i.e., MAQC-III) project, different RNA-Seq platforms were tested across more than ten sites using well-established reference RNA samples with built-in truths in order to assess the discovery and expression-profiling performance of platforms and analysis pipelines. The results demonstrate that novel exon-exon junctions can still be discovered beyond existing comprehensive annotations and sequencing depth. With various investigations encompassing diverse performance metrics (accuracy, precision, reproducibility, mutual information, titration consistency, and mixing ratio recovery) and comparisons to qPCR and microarray platforms, we found high levels of inter-site and cross-platform concordance for differentially expressed genes. However, performance is clearly platform and pipeline dependent, and transcript-level profiling shows larger variation. The SEQC data sets with over 100 billion reads (10 TB) represent the deepest characterization of the transcriptome for any samples and thus provide a unique resource for testing future developments of RNA-Seq in clinical and regulatory settings.

1600T
**Evaluation of Ion Torrent-based rapid deep sequencing for mutation discovery and prevalence screening in rare human myeloproliferative neoplasms and brain tumors**
L. Wang, S. Yamanouchi, L. Lewis, K. Helder, K. Chang, K. Walker, H. Dinh, H. Doddpapeni, D. Muzny, R. Gibbs, C. Lau, D. Wheeler. 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Texas Children’s Cancer and Hematology Centers, Baylor College of Medicine, Houston, TX.

Tumors usually consist of a mixture of normal and tumor cells and most tumors, if not all, are heterogeneous, containing more than one clonal type. Sensitive and accurate detection of somatic mutations in low-purity and heterogeneous tumors is a challenging problem. Deep sequencing is needed to identify mutations with low allele fraction and to present each clonal type properly. For some rare tumors such as myeloid disease and brain tumors, which only a small amount of DNA can be collected, most sequencing technologies are not applicable. Ion AmpliSeq technology, based on ultra-high-multiplex PCR, requires as little as 10 ng of input DNA to target sets of interesting genes, delivers fast library construction and deep coverage on targeted regions, making it ideal for mutation discovery and prevalence screening in these rare, heterogeneous and polyclonal tumors. Here, we selected 40 and 23 interesting genes, respectively, from the discovery studies of myeloid disease and brain tumors, and designed two Ion AmpliSeq arrays targeting the complete coding sequences of interesting genes. A total of 148 samples were sequenced (16 reaction pools/run) on the PGM sequencer and 30 of them were also sequenced on the Proton Sequencers. On average, 98.7 percent of complete coding sequences of targeted genes were successfully covered and a 891x coverage was achieved by PGM run. The Proton run produced about 3 times higher coverage than that of PGM. The variant allelic fraction of mutations with detection rate > 0.01. We identified JAK2-V617F mutations in tumors that had a negative result for whole exome sequencing method. Interestingly, we identified novel recurrent somatic mutations in rare brain tumors including CBL, encoding the ubiquitin E3 ligase that function as a negative regulator of receptor protein tyrosine kinases including KIT, and BCO1L, a transcriptional corepressor and tumor suppressor. We have systematically compared the performance of AmpliSeq PGM and Proton Sequencers including the throughput, on-target coverage, sensitivity and accuracy for substitutions and indels detection and will present here.
1602W
Comparative analysis of six splice-aware alignment and two differential expression assessment tools for RNA-Seq data. J.H. Kim1, O. Evgrafov1,2, K. Wang1,2,3,4, K. Knowles1,2, D. Rabinowitz1, L.B. Jorde1,2, L.T. Boerwinkle2, J. Knowles1,2, 1) Zilkha Neurogenetic Institute, University of Southern California, CA; 2) Psychiatry & the Behavioral Sciences, University of Southern California, CA; 3) Preventive Medicine, University of Southern California, CA.

Many bioinformatics software tools have been developed to perform splice-aware alignments, quantify gene expression levels, or test for differential expression for RNA-Seq data. Despite the wide application of RNA-Seq technologies, it is not well established what are the relative advantages and disadvantages between different analytical approaches. To address this issue, we compared five commonly used alignment tools including TopHat, RNA-STAR, MapSplice, RUM, GSNAP and Perm, and compared two commonly used tools including DESeq and Cufflinks for detecting differential expression levels. The first data set included ~81 million 100bp paired-end reads from a tumor sample. Although four of the tools align more than 97% of reads, TopHat shows a map ratio of only 74%. Speed-wise, RNA-STAR outperformed another with 12 minutes of runtime, while other tools require at least 4 hours in a machine with 12 Intel Xeon 2.66GHz cores. However, RNA-STAR requires more than 23Gb of memory, whereas other tools can run with less than 8Gb memory. We quantitated the gene expression level from each method, compared them to each other and to illuminina HumanHT-12 microarray expression data. Expression data from all tools show more than 0.94 of Spearman’s rank correlation between each other, and all have lower correlation with the microarray data (~0.74). The second data set included ~70 million 91-bp pair-end reads on neural stem cells with and without NRXN1 knockdown. We applied DESeq and Cufflinks to evaluate the p-values of 5 alignments identified by the tool of each method, and found that both methods showed that all 6 aligners identified knockdown of NRXN1 (P=5.0E-05 to 1.5E-05 and 3.5E-04 to 6.9E-04, respectively), and demonstrated that GAPDH has no change in gene expression (P=0.4-0.6 and 0.5-0.6, respectively). However, DESeq identified 54 differentially expressed genes from alignments generated by GSNAP, and 78–120 genes from alignments generated by other 5 aligners, yet Cuffdiff identified 28–47 genes from alignments generated by all aligners, under the same FDR<5% and log2(Fold change)>1 threshold. Nevertheless, the higher number of diversified genes may support the similar correlation to microarray data, but RNA-STAR has a clear speed advantage. Our results may provide a practical guide to readers for selecting aligners, quantifying expression levels and discovering novel splice forms from RNA-Seq data sets.

1603T

High quality and high accuracy are characteristic of Sanger re-sequencing projects and are primary reasons that next generation sequencing projects compliment their results by capillary electrophoresis data validation. We have developed an on-line tool called Primer Designer™ to streamline the NGS-to-Sanger sequencing workflow by taking the laborious task of PCR primer design out of the hands of the researcher by providing pre-designed assays for the human exome. The primer design tool has been created to enable scientists using next generation sequencing to quickly confirm variants discovered in their work by providing the means to quickly search, order and receive suitable pre-designed PCR primers for Sanger sequencing. Using the Primer Designer™ tool to design M13-tailed and non-tailed PCR primers for Sanger sequencing we will demonstrate validation of 28-variants across 24-amplicons and 19-genes using the BDD, BDTv1.1 and BDTv3.1 sequencing chemistries on the 3500xl Genetic Analyzer capillary electrophoresis platform.

1604F
Mobile element scanning (ME-Scan) for Alu insertions in families and populations. D.J. Witherspoon1, W.S. Watkins1, M.A. Batzer2, L.B. Jorde1, 1) Dept of Biological Sciences, Louisiana State University, Baton Rouge, LA.

The human genome carries more than one million copies of the Alu retrotranspon family, the remains of more than 50 million years of activity in our ancestors. Alu retrotransposition and ectopic recombination between Alu copies still generate novel mutations and cause genetic diseases today. A clear understanding of their genomic impact and evolution requires detailed yet comprehensive data on their activity. This is challenging due to the large background of nearly identical sequences in the genome and the modest rate at which new insertions appear in the germline (estimated at 1 per 20 births.) We have developed a targeted high-throughput sequencing technique to efficiently identify novel Alu insertions (Mobile Element Scanning, ME-Scan; Witherspoon et al. BMC Genomics 2010. Witherspoon et al. Genome Res 2013). We have redesigned it to exploit the longer reads and higher capacity of the Illumina HiSeq 2500 platform using standard sequencing parameters. Up to 53 samples can be pooled in a single sequencing library using pairs of custom six-bp indexes. We have applied this design to 290 individuals from diverse populations (HapMap YRI, JPT, TSI, CHB and CEU; Indian Madiga, Mala, Reili and Brahmin; African Alur, Hema, Nande, Luhya and Pygmy) and 85 individuals from 10 parent-offspring pedigrees. The rate of Mendelian inheritance errors (insertions observed in offspring but absent in parents) is <0.5%; no high-confidence de novo Alu insertions have been detected yet. ME-Scan assays of an additional 300 offspring and their parents are underway. Deeper multiplexing and higher sequencing throughput has decreased cost while increasing the coverage of (at Human Genetics, University of Utah, Salt Lake City, UT) 2) Dept of Biological Sciences, Louisiana State University, Baton Rouge, LA.

1605W
Integrative analysis of metabolomics and transcriptomics data: A unified model framework to identify underlying system pathways. C. Ekstrom1, K. Brink-Jensen2. 1) Biostatistics, University of Copenhagen, Copenhagen, Denmark; 2) Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark.

The abundance of high-dimensional measurements in the form of gene expression and mass spectroscopy calls for models to elucidate the underlying biological system. We propose a statistical method that is applicable to dataset consisting of Liquid Chromatography-Mass Spectroscopy (LC-MS) and gene expression (DNA microarray) measurements from the same samples, to identify genes controlling the production of metabolites. Due to the high dimensionality of both LC-MS and DNA microarray data, dimension reduction and variable selection are key elements of the analysis. Our proposed approach starts by identifying the basis functions ("building blocks") that constitute the output from a mass spectrometry experiment. Subsequently, the weights of these basis functions are related to the observations from the corresponding gene expression data in order to identify which genes are associated with specific patterns seen in the metabolite data. The modeling framework is extremely flexible as well as computationally fast and can accommodate treatment effects and other variables related to the experimental design. We demonstrate that within the proposed framework, genes regulating the production of specific metabolites can be identified correctly unless the variation in the noise is more than twice that of the signal.

Modern large-scale human genetics experimental designs take advantage of new technologies that allow for inexpensive high-throughput genotyping to measure genetic variations in hundreds of thousands of individuals. These strategies aim to identify variants in genes associated with important phenotypes. Such unbiased experimental strategies have both benefits, e.g. the opportunity for discovery, but also drawbacks associated with multiple testing considerations. To address these limitations, we developed a method to integrate thousands of publicly available datasets, including gene-expression, protein-protein interaction, and transcription factor binding experiments into networks of tissue-specific functional relationships. We have generated a globally integrated network, as well as networks specific to functional relationships in individual tissues and cell lineages. Such tissue-specific networks encode pathway and functional information in the context of a single tissue or cell lineage. We evaluate the ability of these networks to predict genes associated with disease by using known disease genes from OMIM as seed genes for a state of the art machine learning algorithm. We assess this strategy’s ability to predict the replicating discoveries from 198 genome-wide association studies corresponding to 23 diseases. Our evaluation compares the ability to identify genes containing replicated variants over a set of genes matched by chromosomal region. The state of the art learning algorithm is able to effectively use functional relationship networks to prioritize such genes. Furthermore, we observe that tissue-specific networks are better able to identify replicated discoveries than a global network of functional relationships. Such networks also provide the opportunity to prioritize candidates for epistasis analysis by large-scale molecular experiments. We anticipate that this approach will have broad applications for analysis of both main and epistatic effects, including candidate selection for replication and interpretation of overall findings in a molecular context.


To bring omics data from benchtop to point of care, labs and clinics must be able to handle three types of data with very different properties and requirements. The first, biomedical knowledge, is highly complex, continually evolving and comprises millions of concepts and relationships. The second, medical data including clinical health and outcomes records, is temporal, unstructured, and hard to access. The third, omics data such as whole-genome sequence, is structured but voluminous. Attempts to bridge the three have had limited success. No single data architecture allows efficient querying of these types of data. The lack of scalable infrastructure that can integrate complex biomedical knowledge, temporal medical data, and omics data is a barrier to widespread use of omics data in clinical decisionmaking. Syapse has built a data platform that enables capture, modeling, and query of all three data types, along with applications and programming interfaces for seamless integration with lab and clinical workflows. Using a proprietary, semantic layer based on Resource Description Framework (RDF) and related standards, the Syapse platform enables assembly of a dynamic knowledgebase from biomedical ontologies such as SNOMED CT and OMIM as well as customers’ internally curated knowledge. Similarly, HL7-formatted medical data can be imported and represented as RDF objects. Lastly, Syapse enables federated queries that associate RDF-represented knowledge with omics data while retaining the benefits of scalable storage and indexing techniques. Biologists and clinicians can access the platform through a rich web application layer that enables role-specific customization at any point in the clinical omics workflow. We will describe how biologists and clinicians use Syapse as the infrastructure of an omics learning healthcare system. Syapse supports data mining, query by disease, query by phenotype, and many of the challenges surrounding data analysis and storage volumes are actively being managed and/or solved, the issues associated with integrating and interpreting high-dimensional clinical data, patient sample processing and analysis information, and downstream diagnostics result delivery are still widely unsolved. In this presentation we discuss the issues of data and patient privacy, multisystem integration touch points, data integration and exchange standards, and present examples of solutions from a variety of laboratories that may be useful to organizations looking to implement a NGS data ecosystem to support diagnostic NGS.


In a relatively short period of time next-generation sequencing has established itself as a clinically informative means to pinpoint otherwise difficult to diagnose disease. While the price of sequencing itself has plummeted and many of the challenges surrounding data analysis and storage volumes are actively being managed and/or solved, the issues associated with integrating and interpreting high-dimensional clinical data, patient sample processing and analysis information, and downstream diagnostics result delivery are still widely unsolved. In this presentation we discuss the issues of data and patient privacy, multisystem integration touch points, data integration and exchange standards, and provide examples of solutions from a variety of laboratories that may be useful to organizations looking to implement a NGS data ecosystem to support diagnostic NGS.

1609T From big data to smart data: an open-source solution for genome-scale variant data warehousing and discovery. M.J. Italia1, B. Ruth1, M. Sarmady1, JC. Penn1, D. Naegely1, A. Santani1, M. Dulk1, NB. Spinner1, ID. Krantz4, JW. Pennington1, PS. White1, 4) Center for Biomedical Informatics, The Children’s Hospital Of Philadelphia, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, The Children’s Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA.

NGS-enabled genome-scale diagnostics and research are generating millions of genomic variants on a routine basis. The growth of this data provides the opportunity to derive information content such as local population allele frequencies and variant differentials between samples, patients, and cohorts. However, NGS technologies also present data management and use challenges that overwhelm traditional data and laboratory workflows. We sought to mitigate these challenges and to advance opportunities by developing Varify, a freely-available, open-source integrated variant data warehouse, knowledge base, and analysis suite. We constructed Varify in collaboration with the NHGRI Clinical Sequencing Exploratory Research program and our clinical laboratories. The Varify application was built using Harvest (harvest.research.chop.edu), an open-source, web-based biomedical application development framework. The data warehouse was deployed on the open source PostgreSQL relational database. Unique variants are combined from all patients into one table. This reduces storage requirements while securely linking variants to patients and patient cohorts. Annotation data is structured but voluminous. To bring omics data from benchtop to point of care, labs and clinics must be able to handle three types of data with very different properties and requirements. The first, biomedical knowledge, is highly complex, continually evolving and comprises millions of concepts and relationships. The second, medical data including clinical health and outcomes records, is temporal, unstructured, and hard to access. The third, omics data such as whole-genome sequence, is structured but voluminous. Attempts to bridge the three have had limited success. No single data architecture allows efficient querying of these types of data. The lack of scalable infrastructure that can integrate complex biomedical knowledge, temporal medical data, and omics data is a barrier to widespread use of omics data in clinical decisionmaking. Syapse has built a data platform that enables capture, modeling, and query of all three data types, along with applications and programming interfaces for seamless integration with lab and clinical workflows. Using a proprietary, semantic layer based on Resource Description Framework (RDF) and related standards, the Syapse platform enables assembly of a dynamic knowledgebase from biomedical ontologies such as SNOMED CT and OMIM as well as customers’ internally curated knowledge. Similarly, HL7-formatted medical data can be imported and represented as RDF objects. Lastly, Syapse enables federated queries that associate RDF-represented knowledge with omics data while retaining the benefits of scalable storage and indexing techniques. Biologists and clinicians can access the platform through a rich web application layer that enables role-specific customization at any point in the clinical omics workflow. We will describe how biologists and clinicians use Syapse as the infrastructure of an omics learning healthcare system. Syapse supports data mining, query by disease, query by phenotype, and many of the challenges surrounding data analysis and storage volumes are actively being managed and/or solved, the issues associated with integrating and interpreting high-dimensional clinical data, patient sample processing and analysis information, and downstream diagnostics result delivery are still widely unsolved. In this presentation we discuss the issues of data and patient privacy, multisystem integration touch points, data integration and exchange standards, and provide examples of solutions from a variety of laboratories that may be useful to organizations looking to implement a NGS data ecosystem to support diagnostic NGS.


In a relatively short period of time next-generation sequencing has established itself as a clinically informative means to pinpoint otherwise difficult to diagnose disease. While the price of sequencing itself has plummeted and many of the challenges surrounding data analysis and storage volumes are actively being managed and/or solved, the issues associated with integrating and interpreting high-dimensional clinical data, patient sample processing and analysis information, and downstream diagnostics result delivery are still widely unsolved. In this presentation we discuss the issues of data and patient privacy, multisystem integration touch points, data integration and exchange standards, and provide examples of solutions from a variety of laboratories that may be useful to organizations looking to implement a NGS data ecosystem to support diagnostic NGS.

1609T From big data to smart data: an open-source solution for genome-scale variant data warehousing and discovery. M.J. Italia1, B. Ruth1, M. Sarmady1, JC. Penn1, D. Naegely1, A. Santani1, M. Dulk1, NB. Spinner1, ID. Krantz4, JW. Pennington1, PS. White1, 4) Center for Biomedical Informatics, The Children’s Hospital Of Philadelphia, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, The Children’s Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA.

NGS-enabled genome-scale diagnostics and research are generating millions of genomic variants on a routine basis. The growth of this data provides the opportunity to derive information content such as local population allele frequencies and variant differentials between samples, patients, and cohorts. However, NGS technologies also present data management and use challenges that overwhelm traditional data and laboratory workflows. We sought to mitigate these challenges and to advance opportunities by developing Varify, a freely-available, open-source integrated variant data warehouse, knowledge base, and analysis suite. We constructed Varify in collaboration with the NHGRI Clinical Sequencing Exploratory Research program and our clinical laboratories. The Varify application was built using Harvest (harvest.research.chop.edu), an open-source, web-based biomedical application development framework. The data warehouse was deployed on the open source PostgreSQL relational database. Unique variants are combined from all patients into one table. This reduces storage requirements while securely linking variants to patients and patient cohorts. Annotation data include RefSeq transcripts, SIFT, PolyPhen, HGVS nomenclature, HGMD, Alamut, and population frequencies from 1000 genomes & EVS. A variant loading pipeline reads .vcf files annotated by snpEff and links variants to annotations based on genomic coordinates and genotypes. We have scaled Varify to store and deliver interactive real-time filtering on over 100 million unique variants across several thousand patients from exome and whole genome sequencing. Initial capabilities enabled by Varify include patient-specific query and filtration of variants using complex annotation criteria; calculation of allele frequency for custom cohorts; capture of analyst decisions and evidence on pathogenicity; and ad-hoc query of variants across the patients based on gene, phenotype, and clinical characteristics. Varify is currently deployed in a CAP-compliant clinical diagnostic environment, an enterprise research instance, and as a consortium resource for NICHD’s Newborn Screening Translational Research Network. Varify has added value in both the clinical diagnostic and research setting by transitioning from stand-alone, partially annotated variant files into an integrated resource that adds significant new value to diagnostic and discovery projects.
1610F Integration of ENCODE datasets to epigenomics analysis using colored 
De Bruijn graphs. C. Joly Beaufuant, J. Corbelli, A. Droll. CHU de Quebec, 
Université Laval, Quebec City, Quebec, Canada.

Recently, the ENCODE consortium produced a wealth of genomic data in 
order to start cataloging functional elements in the human genome. The 
challenge is now to integrate this information in our day-to-day analysis of 
smaller scale experiments to better understand the biological data. This is 
not a trivial task considering the difficulty to handle all this information and 
the sheer size of the datasets produced. We propose to use colored De Bruijn 
graphs as a mean to quickly identify genomic regions from epigenomics 
experiments that overlap regions previously described by the ENCODE 
project. This data structure that has successfully been implemented in multi-
ple de novo assemblers can be easily adapted to work with genomic regions 
instead of raw sequencing data. The elements in the graph can then be 
colored using external datasets, like those produced by ENCODE. The 
contigs can then be sorted based on their colors to quickly spot specific 
combination of genomic elements that overlap. A prototype for this algorithm 
has been developed in Python to demonstrate the feasibility of the project.
Every using an interpreted programming language, multiples samples can 
be compared in a relatively short amount of times (in a few hours). The 
algorithm is general enough to be able to use any list of sequences for 
the coloring of the graph. The next development step is to implement the algo-
rum as a plugin of the Ray software, a massively distributed genome 
assembler. The distributed nature of the project will give the scientific com-
munity a tool to manage and integrate the flow of data provided by the 
ENCODE project and other next-generation datasets.

1611W The Japanese Reference Genome in Human Variation Database. A. 
Koike1, M. Kawashima2, Y. Suzuki2, H. Sawai2, M. Yoshida1, N. Nishida3, 
Univ. Tokyo.
The HapMap Project and 1000 genome project have revealed genetic 
differences among populations. Building a reference genome in each popula-
tion is a crucial step for exploitation of disease-related variations. We have 
estimated a Japanese reference genome using whole genome sequencing 
data of five healthy Japanese samples and Japanese samples of 1000 genome 
gene, P.J, and Affymetrix 6.0, AXIOM, and Omni-2.5 data of more than 
400 healthy Japanese samples. The SNPs and small deletions/insertions, 
and CNVs are calculated after mapping all reads on the hg19 reference 
gene. The unmapped reads and clipped reads are assembled and inser-
tion positions are identified using clipped reads whose partial sequences 
are mapped on the h919 genome and rest parts of sequences are mapped 
on the assembled sequences. The variation concordances between different 
platforms are higher than 99.9%. The built reference genome is provided 
as one of reference genomes of Human Variation Database in Japanese 
database project.
The Human Variation Database (http://gwas.biosciencedbc.jp/cgi-bin/ 
hvdb/hv_top.cgi) is a repository database to achieve continuous and inten-
sive management of GWAS/NGS data and data-sharing among researchers. 
This database also accumulates variations extracted from scientific litera-
tures by manual curation to systematize the disease-related variations and 
important our understanding of disease mechanisms. Disease-related 20,000 
variation entries are currently registered in the genome level. Furthermore, 
protein-interaction networks with variations can be drawn by a graphic viewer 
to facilitate understanding of synergetic effects between multiple variations in 
the same disease and relationships between similar diseases. In this 
presentation, we overview the influence of calculation method on the Japa-
nese reference genome and population differences of disease-related vari-
tions and phenotype-unrelated variations. Acknowledgement: This work was 
supported by the contract research fund "Database Integration Program" from 
the Japan Science and Technology Agency.

1612T IScore and ALT_rate: Inerring human diversity using genome-wide gene-
based patterns of nucleotide substitution, insertion and deletion. Y. K. 
Chen1, C. Y. Shen1, C. H. Chen1. 1) Genomic Research Center, Academia 
Sinica, 128 Academy Road, Section 2, Nankang, Taipei, Taiwan 115; 2) 
Institute of Biomedical Sciences, Academia Sinica, 128 Academia Road, 
Section 2, Nankang, Taipei, Taiwan 115.

As the technologies of the next-generation sequencing are being devel-
oped, large scale of genomic researches, such as the 1000 Genome Project, 
have carried out using either whole-genome and/or exome sequencing to 
study human genome variation. Most human genome studies have identified 
single nucleotide polymorphisms (SNPs), insertions/deletions (indels), copy number varia-
tions (CNVs), repeat sequences, and rearrangements in the human genome. 
These variants may have well preserved genetic evidences of human diver-
sity among populations. Previous studies have demonstrated that the genome-wide patterns of 
SNPs can successfully infer human diversity at the population level and 
ancestry information at the individual level. However, such an application 
with the density of SNPs and the indels has still remained unclear. In this 
study, we propose two gene-based density ratios being able to result in 
genome-wide patterns for inferring population diversity and individual ances-
try: IScore, the ratio of the number of indels to the number of SNPs in each of 
the UTR and protein coding regions across the human genome, and ALT_rate, 
an alternative measure of the variant density, the total number of the indels and SNPs in each region. We have calculated IScore and ALT_rate on 
the basis of the sequencing data of the 1000 Genome Project. In order to 
efficiently perform computational analyzing and graphical visualization, 
we have innovated and extended the functionality of the existing public open 
resources platform and Treemap. Our results show that the distribution of 
IScore varies among genes across the human genome, but with similar trend 
among the continental groups at most genes. In particular, some genes 
show relative higher IScore in all ethnicity groups, which suggests that some 
segments of the human genome are under stronger selective pressure and some 
prefer over SNPs over Indels. Further clustering analysis identifies genes with 
different distributions of IScore among ethnicity groups and hence differen-
tiates these ethnicity groups. Similar results were found using the alternative 
ALT_rate. We also find that RAT_ratio might be biasedly over-estimat-
ed in some regions with much lower depth/coverage of sequencing and 
they should be corrected before analysis. In summary, these two gene-
based density ratios, IScore and ALT_rate, can be utilized for inferring human 
diversity and individual ancestry. The ratios may be useful to screen the 
sequence of polymorphism patterns in both patient populations and normal populations.

1613F Using ontologies to enhance integration and analysis of ENCODE 
data. V.S. Malladi1, J.S. Stratton1, D.T. Erickson1, E.T. Chan1, E.L. Hong1, 
G. Barber2, G. Binkley3, J. Garcia4, B.C Hitz2, D. Karolchik5, K. Learned2, 
B. Lee6, S. Myasato7, G. Moro8, G.R. Roe9, K. Rosenbloom10, L.D. Rowe11, 
N.R. Podduturu12, M. Simson13, C.A. Sloan1, E. Weiler14, W.J. Kent15, J.M. 
Cherry16, M. Church16, V.S. Malladi1. 1) Stanford University, Department of Genetics, 
2) Center for Biomolecular Science and Engineering, University of California, 
Santa Cruz, CA, 95064.
The Encyclopedia of DNA Elements (ENCODE), Roadmap Epigenomics 
(RMC), and modENCODE are ICGC projects that aim to provide public re-
sources for the scientific community. The goal of ENCODE, now in its 8th year, is to create a comprehensive catalog of functional elements in the human and mouse genomes. The modENCODE 
project shares this goal but focuses its investigation on the model organisms 
C. elegans and D. melanogaster, creating a comparative resource to provide 
insight into human processes. REMC shares similar assays and tissues to 
ENCODE while investigating the human epigenome. Though each project 
has different goals, these projects complement the data generated by the 
ENCODE project.
The ENCODE Data Coordination Center (DCC), which collects all data and 
metadata generated by the ENCODE project, is currently integrating metadata from modENCODE and REMC. To further analyze the 
information that can be performed within and across these three projects, the DCC 
has made use of ontologies to annotate these metadata. The DCC has used 
The Ontology for Biomedical Investigations (OBI:http://obi-ontology.org) 
to facilitate data identification from various assays sharing similar biological 
objectives (e.g. RRBS, MRE-seq), allow investigators to retrieve all data across the three projects matching various assays 
(e. g. RRBS, MRE-seq). Data identification across shared anatomy, morphol-
y and development are accomplished by using and developing cross-
references with between Cell Type Ontology (http://celltypeontology.info/) and 
UBERON (http://uberon.org/). Researchers querying for a biological system 
(e.g. digestive system) will retrieve data generated from tissues and cells 
that comprise that system (e.g. colon and epithelial cell of stomach). Here 
we describe our implementation of ontologies to integrate these projects and 
how it may be used to identify experiments that match the interests of a 
researcher. Data from the ENCODE project can be accessed via the 
ENCODE portal (http://www.encodeproject.org) and the UCSC Genome 
Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).
1614W

A web-based framework for querying genomic relational databases using SQL. S.F. Sacccone, P.L. Jones. Department of Psychiatry, Washington University, St. Louis, MO.

Genomic databases are an important resource for the study of human genetics and genomic medicine. These databases are commonly accessed through a wealth of visual tools such as genome browsers and other graphical web-based query interfaces. Our BioQ web application (http://bioq.saclab.net) is an SQL-based query tool that provides a flexible framework for querying a given list of heterogeneous genomic features such as chromosomal segments, genes, transcripts, exons and variants across a number of common genomic databases such as the 1000 Genomes Project, dbSNP, Ensembl, VEGA and the European Nucleotide Archive. We introduce new tools in BioQ that allow users to build SQL queries while utilizing various relationships between genomic features such as genes within genomic regions, transcripts associated with genes, exons associated with transcripts and genetic variants within exons. These relationships and their order of execution can be configured by the user in the BioQ web interface in order to define powerful SQL queries across multiple heterogeneous genomic databases. BioQ provides extensive interactive documentation and visualization tools that help the user understand how schemas from different heterogeneous databases are mapped to one another in order to implement genomic feature interrelationships as well as detailed information on data provenance and experimental process flow for data from individual databases. We will demonstrate how this method can be applied to common data integration tasks such as the annotation of results from whole-genome genetic association studies.

1615T


Nucleic acid sequencing is becoming ubiquitous, with approximately 2,500 high-throughput (Hi-Seq & Proton) instruments, and 3,600 mid-throughput (PGM and MiSeq) instruments placed in labs around the world. The bottleneck is now longer biochemical; we are now able to rapidly generate large amounts of sequence data that is high-quality and relatively low cost. The bottleneck is analysis. A well-powered study involves hundreds or thousands of individual samples, and requires a multi-dimensional analysis that incorporates transcript levels, mutations, structural changes, as well as clinical information such as time-to-event or magnitude of disease. Robust statistical methods are required that understand concepts such as gene regions, pathways, and functional classifications. Just as important is the ability to access and manage projects, and communicate results. Station X has developed a software environment called GenePool™ for the management, analysis and communication of genomic information for cohort-scale biomarker studies. GenePool has been architected to support thousands of whole genomes, exomes, and transcriptomes in either a cloud environment or stored locally. Best-practice statistical analyses methods are available in a context-specific drop-down menu. GenePool is designed to return results in seconds rather than minutes or hours, which enables users to iterate their analyses quickly. Data can be filtered based on statistical ranges, functional categories, and clinical meta-data. Station X has also developed dynamic visualizations that can communicate large data sets and their results in a single view. Collaboration is facilitated through shared permissions, rather than maintaining multiple copies of the same dataset. One of the most valuable public data resources is the Tumor Cancer Genome Atlas. However, it can be difficult to access and the clinical meta-data is difficult to incorporate. Station X has integrated and curated the entire TCGA dataset and has identified highly significant genomic biomarkers in breast and lung cancers that stratify patients based on disease outcome. Two examples will be shown. GenePool enables rapid genomic biomarker discovery through best-practice analyses accessed via an easy-to-use user interface, management of cohort-scale data sets, and communication of results through visualization and shared data. Highly significant biomarkers of patient outcome have been identified through analysis of the TCGA data.

1616F


Life science researchers are continually using ways to leverage cloud computing, and applications now range from web-based apps to enterprise wide big data/compute. Optimal performance of diverse applications requires a broad set of cloud topologies and developers are often finding themselves locked into clouds that fail to meet their computing needs or storage demands. Inter-cloud porting would seem the answer and vendors now provide cloud portability services, however most are not specialized to the needs of the life science research or require interim migration steps to run applications on each different cloud. Few provide the researcher with dashboards and a benchmarking system to monitor performance across clouds and researchers again end up in a sub-optimal performance environment. CiiQr has developed a new method, which has now been applied to the scientific computing industry, which offers both straightforward app migration and cloud portability. Our system provides dynamic performance monitoring for the applications researchers use in the cloud and minimizes the development timeframe, with many standard components such as Python, R, and Perl already available and optimized. We will present the results of a real-world case study where one such life science company has benefited from the CiiQr deployment of their packages. The two packages in question serve similar components of the medical industry yet have significantly varying infrastructure demands. Both applications required stringent security (HIPAA Compliance) and the ability to deploy as Software as a Service (SaaS). The first package needed to be 24/7 accessible with a fast customer interface at a single institution, but had low bandwidth, storage and computation needs; while the second required scaling to huge storage, high bandwidth, global access and moderate computational needs. We will demonstrate how porting to the cloud using the CiiQr platform required less effort than performing the implementation across multiple cloud providers, reducing development costs and time to deployment while ensuring maximized performance for the CiiQr Cloud. The two packages were methodically tested for which cloud provided the best performance for each program within hours. We will summarize the general cases as seen by our customers and present guidelines for choosing the best nodes within a cloud, and the best cloud provider, for the most common types of applications.

1617W


The next generation of medicine is envisioned to specifically tailor treatments for patients based on their unique genetic profile and lifestyle. With the rapidly decreasing cost of DNA sequencing and the large investments of medical institutes in digitizing medical records this vision is almost at reach. This new revolution requires the integration of multiple streams of data from different sources - whole patient genome, condition-specific genomic measurements, clinical data from the medical record, and other information including clinical and medical literature - to diagnose diseases, identify relevant treatment options, and monitor response to therapies. This wealth of information must make use of automated analysis to support physicians in making clinical decisions. In order to address these challenges, we recently launched the ‘Clinical Genomic Expert’ an initiative to build an open-source framework that will utilize machine learning algorithms that can incorporate genetic and clinical phenotypes to model the relationships between complex diseases and genome variation, identify a patient’s susceptibility to disease, and predict which therapies might be most effective or cause the fewest side effects. The framework aims to provide an end-to-end solution for medical centers supporting the processing of variant calls as well as data from the Electronic Medical Records (EMR) or patient genomic data. The Clinical Genomic Expert implements several machine learning algorithms that can efficiently solve regression problems on network or tree structures between multiple phenotypes such as time series measurements of traits. These methods account for the inherent structured sparsity in all types of data. The framework also aims to provide a graphical user interface for physicians or clinical researchers to communicate the results of the analysis. Ultimately we hope that the initiative provides a platform for sharing disease models that will allow collaborative work between multiple medical centers without compromising private patient information. URL: http://www.genomic-expert.org/.
1618T
Mining human genetic variation with GEMINI - a novel integrative framework for exploratory analysis. U. Paila1, B. Chapman2, R. Kirchner2, A. Quinan3, 1 Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) School of Public Health, University of Boston, Boston, MA.

Advancements in DNA sequencing technologies have given investigators an unprecedented opportunity to rapidly sequence many genomes for studies of the genetic basis of human disease. High-throughput approaches for human genomes have been instrumental in the identification of new DNA variants linked to disease. However, considering the size and complexity of data generated by such studies, it is challenging to isolate relevant variants without integrating genomic ‘knowledge’ associated with them (e.g. genes, regulatory elements, protein domains, GERP, 1000 Genomes etc.).

Here we introduce GEMINI (Genome MINing), a flexible software tool that addresses these analytical challenges by providing a novel framework for exploratory analysis of human genetic variation in the context of crucial genome annotations (e.g. ENCODE, KEGG, OMIM, dbSNP, 1000 Genomes). GEMINI accommodates studies involving thousands of samples, enables the integration of custom annotations, and supports reproducibility by allowing researchers to ‘version’ their research. By augmenting many large and heterogeneous genome annotations (both coding & non-coding) into a unified database, it provides a powerful tool for data exploration that addresses the hurdles of large-scale disease research. It allows users to compose complex queries based on sample genotypes (e.g. return all loss of function variants that are homozygous for only sample A’), inheritance patterns, and both pre-installed and custom genome annotations to address their specific research needs.

The GEMINI framework facilitates discovery in a broad range of research including studies of the genetic basis of human diseases, personal genomes, unsolved Mendelian disorders, explorations of rare variants in large pedigrees, and genome-wide case-control studies. The portability and flexibility of the tool will allow other developers to leverage GEMINI to create new tools and develop novel approaches to prioritize genetic variation in diverse contexts. We foresee GEMINI as an outstanding future resource for medical and cancer genomics, giving the depth and weight of annotations it carries, as well as the ease and speed of data development (e.g. methods for identifying significantly mutated genes, collapsing methods, network analysis). GEMINI is a self-contained, open source tool that will benefit the human genetics community.

1620W
The European Variation Archive (EVA) at the EBI. I. Lappalainen, D. Spalding, S. Saha, L. Skipper, J. Ameida-King, V. Kumandur, P. Flichek, J. Paschall, EBIL-EBI Wellcome Trust, Cambridge, United Kingdom.

The European Bioinformatics Institute (EBI) has launched a new genetic variation service called the European Variation Archive (EVA). The EVA will consolidate, accession and provide new ways to access genetic variation data on behalf of the European Genome-phenome Archive (EGA) for data requiring such access. For structural and copy number variants, EVA efforts include continued development of the existing Database of Genomic Variants archive (DGVa) for data requiring such access. The EVA data warehouse layer is under development and based on the NoSQL document centered MongoDB database. This solution manages queries across submitted datasets as well as global data mining functions such as clustering of variants across studies and allele frequency calculation. The use of document based databases as the layer benefits from speed of loading, flexible schema changes to keep up with evolving format standards and dramatically simplifies the data architecture as compared with a relational model. The EVA helpdesk assists submitters or provides more information about the service at eva-helpdesk@ebi.ac.uk.

1621T
RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. R. Thompson1, I. Sutt1, K. Bunby1, E. Heilap1, L. Johnston1, D. Tarascio1, L. Monaco2, C. Beroud2, M. Hansson3, H. Lochmüller1 on behalf of RD-Connect. 1) The MRC Centre for Neuromuscular Diseases at Newcastle, Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, United Kingdom; 2) Centro Nacional de Analisi Genómico, Baldiri Reixac, Barcelona, Spain; 3) Istituto Superiore di Sanità, Viale Regina Elena, Roma, Italy; 4) Fondazione Telethon, Piazza Cavour 1, Milan, Italy; 5) Inserm U491 - Génétique Médicale et Développement, Faculté de Médecine de la Timone 13385 Marseille Cedex 05, France; 6) Centre for Research Ethics & Bioethics, Uppsala University, Box 564, SE-751 22 Uppsala, Sweden.

Despite many examples of excellent practice, rare disease (RD) research is still fragmented among different disease type and disease groups. Individual efforts often have little interoperability and almost no systematic connection of detailed clinical information with genetic information, biomaterial availability or research/trial datasets. Linking data at both an individual-patient and disease level enhances users’ ability to determine the feasibility and clinical utility of the disease of interest, while providing access to data from other research groups in a secure fashion allows researchers in multiple institutions to compare results and gain new insights. Funded by the EU Seventh Framework Programme under the International Rare Diseases Research Consortium (IRDRC), RD-Connect is a global infrastructure project which links databases, registries, biobanks and clinical bioinformatics data used in RD research into a central research resource. RD-Connect’s primary objectives are to provide a centralised hub for disease and development of common standards for RD patient registries by developing a common registry infrastructure and data elements / Harmonisation and development of common standards and catalogue for RD biobanks that collect and provide standardised, quality-controlled biomaterials for translational research / Development of clinical bioinformatics tools for RD research / Development of tools for the rapid discovery of new disease genes, pathways and therapeutic targets / Development of an integrated platform to host and analyse data from omics research projects / Development of mechanisms for incorporating patient interests and engaging stakeholders / Development of a framework for a regulatory framework for linking medical and personal data related to RD-Connect will accept data generated by IRDRC projects such as EUFenOmics, which focuses on causes, diagnostics, biomarkers and disease models for rare kidney disorders, and NeuroMics, which uses next generation whole exome sequencing to increase genetic knowledge of rare neurodegenerative and neuromuscular disorders. The ‘silenced’ nature of rare research efforts is a continued bottleneck for cutting-edge research towards diagnosis and therapy. RD-Connect aims to unite existing infrastructures and integrate the latest tools in order to create a comprehensive combined omics data, biobanking, data analysis and patient registry platform for RD used by researchers across the world.
1622F
A comparison of imputation quality: combining different GWAS platforms. E.P.A van Iperen1,2, G.K.Hovingh3, F.W. Ausselberg3,4, A.H. Zwinderman1. 1) Department of Clinical Epidemiology, Biostatistics and Bio-informatics, Academic Medical Center, Amsterdam, The Netherlands; 2) Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands; 3) Department of Vascular Medicine, Academic Medical Center, The Netherlands; 4) Division of Heart and Lungs, Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands.

Background: In the past decade many Genome-wide Association Studies (GWAS) were performed which have discovered many new associations between single-nucleotide polymorphisms (SNPs) and different phenotypes. Imputation methods are widely used in GWAS. They facilitate the association of ungenotyped variants with phenotypes. Imputation methods can also be used to combine and analyse data genotyped on different genotyping arrays. In this study we investigated if there is a difference in imputation quality between two different approaches of combining GWAS data from different genotyping platforms. We will investigate if combining data from different platforms before the actual imputation performs better than combining the data from different platforms after imputation.

Methods: Existing genotype data from the AMC-PAS cohort were used for this study. The samples were genotyped on three different platforms. A total of 706 individuals were genotyped on the Metabochip (220K SNPs), a total of 757 individuals were genotyped on the 50K gene-centric Human CVD BeatChip (50K SNPs) and a total of 955 individuals was genotyped on the HumanExome chip (240K SNPs). After pre-imputation quality control (QC), Minimac in combination with MaCH was used for the imputation of all samples with the 1000genomes reference panel. All markers with an r2 value of <0.3 were excluded in our post-imputation QC.

Results: All three datasets were carefully matched on strand, SNP ID and genomic coordinates. This resulted in a dataset of 979 unique individuals and a total of unique 258925 markers. Our results suggest that the imputation of the different platforms independently performs slightly better than combining the different platforms before imputation. As an example: In the analysis the three were all imputed separately we imputed successfully 55552 unique SNPs on chromosome 22. In the analysis where the 3 platforms were imputed in combination we achieved a total of unique 50500 SNPs. We observed similar findings for the other chromosomes.

Conclusions: Based on the preliminary results we concluded that combining the data from three different platforms together after imputation performs better than combining the data of the 3 platforms before imputation.

1623W
RNA-Seq optimization with eQTL gold standards. S.E. Ellis1, S. Gupta2, F.N. Ashar1, J.S. Bader1, A.B. West1, D.E. Arking1, Johns Hopkins University School of Medicine, Baltimore, Maryland; 2) University of Alabama School of Medicine, Birmingham, Alabama.

Over the past decade, there has been intense interest in garnering a more complete understanding of genome-wide gene expression, with RNA-Seq becoming the gold standard approach to studying the human transcriptome. Accordingly, RNA-Seq experiments have been optimized for library preparation, mapping, and gene expression estimation. These methods, however, have revealed weaknesses in the next stages of analysis of differential expression, with results sensitive to systematic sample stratification or, in more extreme cases, to outliers. To address these issues, we turn to expression quantitative trait loci (eQTLs), or biologically meaningful loci at which gene expression is modified by genotype, to optimize RNA-Seq data analysis. Here, we propose a method in which we utilize previously published eQTLs at the center of our data analysis framework in combination with DNA genotypes and RNA-Seq expression data to demonstrate our ability to assess not only the integrity of sequencing data but also the appropriateness of data handling procedures. Using postmortem brain samples from autism-affected cases and controls, we demonstrate that our method successfully identifies sample outliers and low levels of sample contamination that, if not otherwise identified, would interfere with accurate gene expression quantification. After outlier detection, we assess data normalization procedures, to demonstrate the biologic value of covariate inclusion in RNA-Seq analyses, support the inclusion of UTRs in gene annotation, and highlight the importance of removing outlier samples on a gene-by-gene basis. These steps allow for more accurate approximation of gene expression levels and can be combined with downstream disease-based comparisons ultimately contributing to understanding of genetic variation, gene expression, and disease.

1624T
A graph-based integration with multi-omics data and genomic knowledge for cancer clinical outcome prediction. D. Kim1, J. Joung2, K. Sohn3, Z. Xu4, A. Rishith5, J.H. Kim1, D. Kim1, Department of Biotechnology, Materials, and Molecular Biology, Pennsylvania State University, University Park, PA, USA; 2) Seoul National University Biomedical Informatics (SNUBI), Div. of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea; 3) Translational Bioinformatics Lab (TBL), Samsung Genome Institute (SGI), Samsung Medical Center, Seoul, Korea; 4) Department of Information and Computer Engineering, Ajou University, Suwon, Korea; 5) Department of Industrial & Information Systems Engineering, Ajou University, San San, Wondchun-dong, Yeoju-gu, 443-749, Suwon, Korea; 6) Systems Biomedical Informatics Research Center, Seoul National University, Seoul 110799, Korea.

Cancer is a complex disease, which can be dysregulated through multiple mechanisms. Therefore, no single level of genomic data fully elucidates tumor behavior since there are many genomic variations within or between levels in a biological system such as copy number variants, DNA methylation, alternative splicing, miRNA expression, and alternative translation, post translational modification, etc. Nowadays, a number of heterogeneous types of data have become more available from the Cancer Genome Atlas (TCGA), generating multiple molecular levels of omics dimensions from genome to phenotype. Given multilevels of genomic data, information from one level to another may lead to some clues that help to uncover an unknown biological knowledge. Thus, integration of different levels of data can aid in extracting new knowledge by drawing an integrative conclusion from many pieces of information collected from diverse types of genomic data. Previously, we have proposed a graph-based framework that integrates with multi-layers of genomic data, copy number alteration, DNA methylation, gene expression, and miRNA expression, for the cancer clinical outcome prediction. Here, we propose a new graph-based framework for integrating different levels of genomic data and genomic knowledge in order to overcome variability of diagnostic or predictive models and to increase their performances. As a pilot task, we used an ovarian cancer dataset from TCGA for the stage, grade, and survival outcomes. Integrating multi-omics data with genomic knowledge to construct pre-defined features results in higher performance in clinical outcomes prediction. For example, the model with gene expression data performed with an AUC of 0.7866. However, the models of the integration with pathway, Gene Ontology (GO), chromosomal gene-set, and motif gene-set consistently outperformed the model with genomic data alone. The model with genomic data alone achieved 0.8254, and 0.8179 respectively. Furthermore, incorporation of genomic knowledge offers more interpretable results from the obtained signatures of pre-defined genomic knowledge and thus provides more insight into the complex molecular mechanisms in cancer. With integration of multi-omics data and genomic knowledge, understanding the molecular pathogenesis and underlying biology in cancer is expected to provide better guidance for improved prognostic indicators and effective therapies.

1625F
Phenotype to Genome: Quantitative Trait Loci (QTL) in the Mouse Genome Informatics (MGI) Database. Y.S. Zhu1, D.J. Reed2, P. Hale3, C.J. Bull1. Mouse Genome Informatics database group. Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME.

The laboratory mouse is an ideal experimental model organism for identifying the genetic basis of complex phenotypes and disease traits. QTL studies in the laboratory mouse have a high degree of concordance with mapped phenotype regions in the human genome but require far fewer resources and time than do mapping studies in human populations. The Mouse Genome Informatics database (MGI; http://www.informatics.jax.org) includes a compendium of over 4,700 published QTL mapping data for the mouse from a variety of sources, including large population and high throughput data. The database includes descriptions of the QTL mapping experiments along with links to the published literature and, in some cases, links to the underlying genotype and phenotype data maintained by the QTL Archive project (http://www.qtlarchive.org/). The genomic coordinates for the markers used in map QTL regions are used to show the genome context of QTLs using MGI’s Mouse Genome Browser. The Genome Browser allows users to explore the all of the genome features annotated in the mapped regions. Newly implemented visualization tools at MGI for navigating regions of conserved synteny between mouse and human facilitate comparative genomics approaches to QTL mapping studies. MGI’s integration of genome features with their biological attributes (phenotype, expression, function) facilitates the use of the database system to support in silico candidate gene analysis.
1626W
Comprehensive analyses of the functional roles of KAO-NASHI genes in the vertebrate organogenesis using medaka model. Y. Tonoyama1,2, A. Shinjizu1, N. Wakakuwa1, Y. Shinjizu1, N. Shinjizu1. 1) Advanced Research for Genome Super Power Center, Keio university; 2) Iwate Tohoku Medical Megabank Organization, Iwate Medical University; 3) Nagahama Institute for Bio-Science and Technology.

The human genome project has provided a computer-estimation on 23,000 protein-coding genes in the human genome. However, many of these protein-coding genes are not fully proven for their existence by experimental evidence. In general, proteins with known motifs are readily classified, but substantial numbers of protein have no obvious motifs in their deduced amino acid sequences. We previously designated these genes/proteins without obvious motifs as KAO-NASHI (Face-less) and initiated a project to unveil their face (kao) by employing comparative genomics and knockdown analysis. We extracted 1,000 KAO-NASHI genes from human genome sequence by step-wise filtration with InterPro motif analysis, BLAST homology search and PubMed document search. A small fish medaka (Oryzias latipes) was chosen as an experimental system to knockdown medaka orthologs of human KAO-NASHI genes with morpholino-antisense oligos (MO). As an initial study, we designed MOs to target translation initiation sites of 100 medaka KAO-NASHI genes. When these MOs were injected into fertilized medaka embryos, their morphogenesis at early developmental stages was disturbed and morphological changes were observed at significant rates. Here, we present the data on a few KAO-NASHI genes in which the MO injected-embryos exhibited significant defective organogenesis involved in brain formation and/or early embryonic vascular development, whereas almost all 5-base mismatch MO-injected embryos exhibited normal phenotypic. Through this process, we identified several novel non-coding genes. Additionally, we performed differential expression analyses between Rx3 homeobox gene-depleted medaka embryos and control embryos using medaka DNA-microarray to seek for KAO-NASHI genes involved in the eye development. As a result, the expression of three KAO-NASHI genes was consistently affected after knockdown of Rx3, which generates ᵂⁿ涿ophthalmia medaka (without eye), implicating that these genes play important roles during development of vertebrate eye. Thus, our approach using medaka model provides the functional information on the human KAO-NASHI genes/proteins during vertebrate organogenesis.

1627T
Rapid and Uniform Whole Exome Libraries from 50 ng of DNA. M. Andersen1, K. Rhodes1, S. Roman1, A. Broomer1, C. Van Looy1, D. Topacio1, M. Allen1, S. Rozenzhak1, G. Liu1. 1) Life Technologies, Carlsbad, CA; 2) Life Technologies, Foster City, CA.

Exome sequencing is a powerful and cost-effective way to research variants linked to genetic disorders. Pairing simple fast exome selection and library generation with the Ion Proton™ sequencer allows two exome libraries to be prepared, sequenced, analyzed, and annotated in two days. This new method allows for the production of whole exome selection and libraries from as little as 50 ng of DNA in 6 hours with only 30 minutes hands-on time. Eight barcoded libraries can be prepared simultaneously on a single plate from 2.4 million uniformly amplified target regions. Genomic DNA of HapMap samples NA12878 (a female genome of European ancestry) and NA19240 (a female genome of Yoruba ancestry) that have been deeply sequenced by the 1000 Genomes Project were used to assess the performance of the Ion Proton™ System with the new exome selection and library method. Rapid read mapping, determination of exome coverage metrics, and variant calling were automatically performed using the Torrent Suite™ Software and Torrent Variant Caller. Variant annotations were completed with Ion Reporter™ software. Greater than 90% of 58 million bases in each exome were covered at 20X and >99% covered at 1X, with >98% concordance with known SNPs.

1628F
ZoomMiR, a computational method to predict and screen for variants that disrupt microRNA binding and activity. A.N. Dubinsky1,2, L. Edsall1,2, A.R. La Spada1,2,3, T. Gaasterland1,2. The NEIGHBOR Consortium for Regenerative Medicine, La Jolla, CA; 3) Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA.

Purpose: We developed a computer program and assay procedure to detect and validate genome variants that alter microRNA binding sites on spliced RNA transcripts through high-sensitivity computational and scalable functional assay screens. Background: Whole genome and exome sequencing identify many variants in untranslated regions (UTR) and introns near exons, i.e., on the order of 25,000 - 30,000 per exome. MicroRNAs bind to RNA transcripts predominantly in the 3'UTR and, with some frequency, to coding sequences and 5'UTR. With the abundance of human exomes and complete genomes sequenced in recent years, tools and technologies with low false positive rates are needed for accurate, efficient, reliable screening of microRNA binding site disruption by genome variants. Methods and Results: Available algorithms and databases for computationally predicted microRNA binding sites have limitations that hinder their application to assess binding site differences due to genome variation. Existing binding site predictions require substantial computation and, in some methods, simulation of binding to generate predictive scores. For example, many databases that serve computed sites neglect the second arm of microRNA hairpins, which have been observed to be loaded preferentially into the microRNA RISC complex. To address this shortcoming we developed a novel, efficient process called ZoomMiR which allows detection of microRNA:miRNA seed sites and ranks potential microRNAs for experimental screens. To validate the computational output demonstrated that it disrupted miRNA seed sites in a reporter gene system. This dual luciferase-based reporter system permits the quantitation of binding of, and hence regulation by, a computationally ranked miRNA. It is sensitive to the effect of the three nucleotides preceding the 2-nucleotide seed site and with a perfect seed site. We applied this method to UTR variations found in 420 exomes from patients with primary open angle glaucoma. Importantly, we analyzed a 3'UTR glaucoma risk SNP in CDKN2B and demonstrated that it disrupted microRNA binding sites and can reveal molecular mechanisms for at-risk-SNPs found through GWAS.

1629W

Human betacoronavirus 2c EMC/2012 was isolated from a subject with acute pneumonia in June 2012. This talk will focus on the characterization of the genome sequence, arrangement, and phylogenetic relation with known coronaviruses sampled from around the world including the Middle East respiratory symptom coronavirus (MERS-CoV). The genome contains 30,119 nucleotides and contains at least 10 predicted open reading frames, 9 of which are predicted to be expressed from a nested set of seven subgenomic mRNAs. Phylogenetic analysis of virulence factors of sequenced genomes indicated that HCoV-EMC/2012 is most closely related to species within the genus Betacoronavirus.
1630T Comparative analysis of whole exome and whole genome DNA sequencing. M.M. Parker, M.A. Taub, K.N. Hetrick, H. Ling, R.A. Mathias, H. Betancur, K. Alibabu-Hejazi, F.P. Eriksson, T.H. Beauty, J.E. Bailey-Wilson, T.H. Beauty. 1) Epidemiology, Johns Hopkins University, Baltimore, MD; 2) Biostatistics, Johns Hopkins University, Baltimore, MD; 3) Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD; 4) Allergy and Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, MD; 5) Ibn Al-Nafees Hospital, Damascus; 6) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 7) National Human Genome Research Institute (NHGRI), Baltimore, MD 21224.

Background: Four subjects were sequenced with both whole exome (WES) and whole genome sequencing (WGS) as part of larger studies to identify causal variants for oral clefts. We assess the effect of exome capture technology on sequencing coverage and discovery of single nucleotide variants (SNV) using these data. Sequencing: WES was performed at the Center for Inherited Disease Research (Baltimore, MD) using Illumina’s HiSeq2000 sequencer, and exome capture was carried out using Agilent’s Human All Exon V3 50 Mb kit following manufacturer’s guidelines. WGS was performed by Illumina (San Diego, CA) using their HiSeq 2000 sequencer. Sequence reads were aligned to the human reference genome hg19, and we performed local re-alignment around indels and base quality score recalibration using GATK (v.2.4). Variant calling was done using GATK’s UnifiedGenotyper. Analyses presented here were restricted to regions on chromosome 20 targeted by Agilent’s capture kit including 5,203 regions totaling 1,246,339 bases. Results: Mean target coverage of the 4 WES samples (93.2x) was higher than the mean coverage of the 4 WGS samples (40.6x), however the variance in coverage of WGS was over 3 times smaller than that of WES. Exon and Genes coverage was highest at 100x and 400 bp. For some degraded DNA samples, the availability of longer fragments will be significantly lower compared to the shorter fragments leading to non-uniform sequencing depths over the targeted regions. We demonstrate that by employing target sequencing by design, as well as targeting both polarities of the genomic DNA molecules, the efficiency of capture is significantly increased leading to higher and more consistent sequencing coverage between samples of different quality. We demonstrate that improved experimental design combined with an increased capture efficiency enable lower DNA input amounts while keeping a high sensitivity for alleles down to 1% in frequency.


Detection of rare DNA mutations from heterogeneous formalin-fixed paraffin embedded (FFPE) tissue samples involves several challenges. The sample amounts retrieved from biopsies are often scarce and the DNA is usually degraded and chemically modified as an effect of the formalin fixation. To maximize the chance of success when performing targeted sequencing studies on such samples, it is imperative to have a well-balanced experimental design as well as high efficiency throughout the enrichment and library preparation in order to ensure the mutations are both sampled and sequenced. The quality of the sample and the expected mutation frequency sets requirements on the minimum amount of input DNA that is needed to ensure that all mutations are sampled. Agilent’s FFPE-Derived DNA Quality Assessment protocol is based on a simple multiplex PCR of different amplicon lengths that provide a fast and reliable estimation of the sample quality. Based on the outcome of this assay, measures can be taken to optimize the experimental design to ensure maximum probability of finding all variants. These measures include determining the minimum input DNA amount and required sequencing depth. Increased enrichment and library preparation efficiency will further help minimizing the required DNA input amount while maximizing the number of DNA molecules available for sequencing thus increasing the limit of detection for rare variants. HaloPlex PCR can be used to enrich for customized target regions and to prepare libraries for sequencing on Illumina or Ion Torrent in less than 8 hours with target regions ranging from a single gene up to the complete exome. A standard HaloPlex assay will cover on average 10,000 targets between 100 and 400 bp. For some degraded DNA samples, the availability of the longer fragments will be significantly lower compared to the shorter fragments leading to non-uniform sequencing depths over the targeted regions. We demonstrate that by employing a target sequencing by design, as well as targeting both polarities of the genomic DNA molecules, the efficiency of capture is significantly increased leading to higher and more consistent sequencing coverage between samples of different quality. We demonstrate that improved experimental design combined with an increased capture efficiency enable lower DNA input amounts while keeping a high sensitivity for alleles down to 1% in frequency.

1633T Globus Genomics: Enabling high-throughput analysis and management of NGS data for neurodevelopmental disorders. D. Sulakhe1, A. Paciorkowski3, G. Mirzaa2, R. Madduri1, Q. Zhang2, K. Aldinger1, J. Bennett1, H. Lacinski1, P. Dave1, W. Dobyns3, 1) Computation Institute, University of Chicago, Chicago, IL; 2) Center for Integrative Brain Research, Seattle Children’s Research Institute, University of Washington, Seattle, WA; 3) Center for Neural Development and Disease, University of Rochester Medical Center, Rochester, NY.

The availability of low-cost high-throughput sequencing methods in the form of next-generation sequencing is revolutionizing translational research. However, handling such large volumes of sequencing datasets and their analyses introduce great challenges including secure and reliable data transfer and storage, as well as many data management and analysis needs. The Globus Genomics platform offers a powerful and efficient tool for the transfer and analysis of nextgen data for clinical and research purposes. The platform hosts an enhanced Galaxy instance with hundreds (539 tools) of widely used next-gen analysis tools and many pre-defined best practices pipelines for whole exome, RNA-Seq, or ChIP-seq data analysis. Unlimited scalability, anonymous analysis of numerous samples, thousands of amplicons of sizes due to the platform’s ability to provision on-demand compute clusters on Amazon and submit workflows to that cluster from Galaxy. Globus Genomics allows dynamic tool specific provisioning of Amazon EC2 nodes, thus accommodating a wide range of CPU and memory intensive analysis including requiring varying compute capabilities that helps in drastically reducing execution times. This platform has been used successfully at the Dobyns laboratory at the University of Washington to transfer hundreds of exomes amounting to tens of terabytes from Perkin-Elmer sequencing centers to Amazon AWS, allowing data to be ready for analysis a few hours after sequence generation. The use of Globus Genomics successfully cut back on data transfer time from a few weeks to a few hours. Furthermore, the platform has provided a 20x performance improvement in the upstream analysis by allowing the simultaneous analysis of 20 exomes in parallel. The Globus Genomics platform offers a powerful and efficient tool for the transfer and analysis of nextgen data for clinical and research purposes.

The massive amount of data generated by next generation DNA sequencing (NGS) systems frequently necessitates sample multiplexing to reduce cost. As a result, sample library construction is often performed in parallel in high-density plate format, thereby increasing the probability of sample cross-contamination. In clinical settings, such contamination can cause important disease-causing mutations to be missed. It is therefore imperative that clinical labs employing sample multiplexing develop strategies that allow contamination to be detected.

We have developed an approach that enables the detection of contaminated samples run through targeted sequencing assays. This approach is based upon the assumption that contamination will skew the allele fractions measured at genomic positions where the contaminated and contaminating samples differ in genotype, and therefore that skewed allele fractions can be used to detect contaminated samples. To implement this approach we have supplemented a set of molecular inversion probes that capture the coding regions of genes associated with recessive genetic disorders with probes designed to capture single nucleotide polymorphisms that have high minor allele frequencies. Thus, for each assayed sample, we measure the allele fractions at a set of commonly variant positions. These measured values are subsequently compared to a set of reference distributions generated from non-contaminated samples to derive a score that indicates the degree to which the observed fractions deviate from the reference distributions.

Simulation and spike-in experiments illustrate that these scores are correlated with contamination levels and therefore can be utilized to flag samples that are contaminated at levels relevant to clinical sequencing. Thus, we conclude that our allele-fraction based system represents a promising new tool that can be utilized to discern contamination and therefore improve the accuracy of targeted sequencing assays.


DNA sequence analysis and genotyping of biological samples using next-generation sequencing (NGS) has been limited by the small amount of sample available. A single cell comprises only one to four copies of genomic DNA, depending on the organism (haploid or diploid organism) and the cell cycle phase. The DNA amount of a single cell ranges from a few femtograms in bacteria to picograms in mammals. However, a thorough analysis of the genome requires a few hundred nanograms up to micrograms of genomic DNA. Consequently, accurate whole genome amplification (WGA) is required for reliable genetic analysis (e.g., NGS) when genomic DNA is limited, as in the case of single cell DNA.

Usually single-cell genomic analysis, and in particular single-cell sequencing, suffers from incomplete or biased genome amplification with missing or underestimated sequence information. To overcome these typical drawbacks, we developed an easy-to-apply single cell WGA method. This method is based on isothermal multiple displacement amplification (MDA) and consists of an innovative lysis and use of an optimized form of the Phi 29 Polymerase. To prove the method’s robustness for single cell amplification, we amplified a variety of human and bacterial single cells, and checked the resulting genome coverage with NGS and qPCR methods. Discussed are experiments on cell-to-cell variations, GC content in comparison to genomic DNA, percentage of genome coverage with respective error rates, and genome-wide real-time PCR analysis. Overall the new method results in the effective lysis of cells, complete DNA denaturation, and reliable amplification of the whole genome of a single cell with high accuracy and minimal amplification bias.

Subcellular Fractionation Proteomics Is An Indispensable Tool for Polypatharmacology Studies for the Identification of Molecular Targets and Transcription Modulator - A Case Study on the Anticancer Auranofin. S. Tian, F.-M. Siu, Y.M.E. Fung, C.-N. Lok, C.-M. Che. Department of Chemical Science, Key Laboratory for the Synthesis of Functional Materials of Chemical Biology of the Institute of Molecular Technology for Drug Discovery and Synthesis, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China.

In the era of polypharmacology, identify mechanisms of action (MOAs) and molecular targets of drug is highly important. Typical proteomics study determines the protein expression changes in the whole cell lysate of the cancer cells upon drug treatments. This method suffers from problems of low coverage and lack of subcellular location information, and renders identification of molecular targets or MOAs difficult. In the present study, we have developed a subcellular fractionation (SCF) based proteomics approach, and applied to study the MOAs of the anticancer Auranofin (AuRF). Compare to that of the typical proteomics study, SCF approach increase the proteome coverage by 4.8-folds: the number of non-redundant proteins identified using LTQ-Orbitrap Mass Spectrometer by this SCF approach is over 2,700, compared to that of 567 proteins in the typical approach. The enriched proteome revealed HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A Reductase) as the potential inhibitory target for AuRF. This finding was verified by enzyme inhibitory assay with IC50 values of 3.5 μM. Using the nuclear sub-proteome, our analysis revealed that p14ARF (gene as CDKN2A) was enriched in 2.5 folds in transcription modulation. It is hypothesized that the up-regulation of p14ARF was associated with E2F-dependent transcription and p53 pathway. This hypothesis was further verified by luciferase reporter system and western blotting assay. Collectively, the subcellular fractionation proteomics is an indispensable tool for polypharmacology studies.

Objective: To match the increased throughput and decreased turnaround time associated with next-generation sequencing (NGS) platforms, it is essential to have an efficient method for library preparation. The Access Array System is capable of generating amplified material from up to 480 genomic regions across 48 unique samples per IFC. The unique design of the Access Array IFC permits efficient DNA amplification in multiplex reactions and enables successful amplification of challenging DNA samples, such as those from FFPE tissues. In this study, we demonstrate the flexibility and performance of the Access Array System in producing libraries from both high quality and FFPE-derived DNA samples, with consistent results on multiple Illumina and Ion Torrent NGS platforms. Methods: 301 primer pairs were designed covering all exons of the genes for BRCA1, BRCA2, TP53, EGFR and MET. These primer pairs were used in multiplex format with the Access Array system to target-enrich samples for NGS. Samples used included high quality genomic cell line DNA (50 ng/sample) and FFPE-preserved lung tissue DNA (with an input of 100 copies of DNA). Results: The libraries prepared were tested on the Illumina (MiSeq, HiSeq) and Ion Torrent PGM platforms. Differences in performance were observed for each of the platforms tested. Results across Illumina platforms were very consistent, with between 95% to 97% of all reads mapping to the human genome, as compared to up to 85% on the PGM. Of the reads mapping to the human genome, 99.8% of the reads consistently mapped to the specific target region in all platforms. On the MiSeq, 85-96% of all assays tested generated in number of reads within five-fold of the average (median 96%, variance 5%) compared to 88-98% (median 96%, variance 4%) on the HiSeq. On the PGM 314 chip, 85-98% of all assays tested generated a number of reads within five-fold of the average (median 96%, variance 6%), compared to 96-98% (median 97%, variance 1%) on the PGM 316 chip. Comparable results were obtained with FFPE-derived DNA samples. Conclusion: These results demonstrate that the Access Array System is capable of generating excellent data regardless of sequencing platform. This system is also a promising technology for use in routine clinical testing due to its robustness to amplification artifacts and the ability to process samples with over 85% of the target regions detected, from as little as 100 DNA copies, comparing very favorably with published reports.

**1641W**

Use of a Targeted Next Generation Sequencing Approach for the Study of a Cardiac Valve Malformation With Complex Polygenic Heritability. E. P. S. Chervitz, 1 A. A. Zender, 1 D. Corsmeier, 1 S. M. Fitzgerald-Butt, 2 D. Newburger, 1 P. White, 3 V. Garg, 3 K. L. McBride, 1 1) Center Cardiovascular & Pulmonary Research, Nationwide Childrens Hospital, Columbus, OH; 2) Center for Microbial Pathogenesis, Nationwide Childrens Hospital, Columbus, OH; 3) Dept Pediatrics, Ohio State University.

Purpose. Bicuspid aortic valve (BAV) is the most common type of heart defect with a population prevalence of 1-2%. BAV is highly heritable with a complex polygenic etiology, however few causative genes are known. Current methods of BAV discovery are limited by the extent of contributory genetic variants. Sanger sequencing is time and labor intensive, while next-generation sequencing (NGS) remains costly for large cohorts and requires extensive bioinformatics processing. Here, we describe a novel approach to targeted multigene sequencing of pooled samples allowing for focusing of a large number of genes relevant to cardiac development in a well-phenotyped cohort in a cost-effective manner. Methods. Candidate genes were selected based on known causal role, mouse models and cardiac developmental biology. A custom capture of 97 candidate genes was designed using the Agilent SureSelect system to probe whole gene intervals. The cohort of 82 unrelated individuals with echo confirmed BAV was combined into 19 pools with a unique overlapping design; a given variant could be attributed to a single individual on the basis of allele content of the variation.

Results. Target capture identified 99 rare exonic variants involving 34 of the 97 candidate genes. BAV calling using an allele frequency threshold of 2.5% compared favorably to the CRISP algorithm. CRISP identified a total of 51 BAV-associated variants for our cohort, while 2 variants were not called by BAV calling. The result achieved in a cost savings of $78,750 as compared to whole exome sequencing of individuals or $509,000 as compared to whole genome sequencing. Conclusion. Targeted capture allowed for decreased bioinformatics processing by focusing only on those gene pathways thought to be pertinent to the disease under investigation. In addition, the pooling design reduced overall sequencing costs. The CRISP algorithm provided a higher sensitivity than BAV in this design. The data suggest a role for targeted NGS of pooled samples when investigating birth defects which are proposed to have a complex and polygenic inheritance.

**1640F**


The current amount of data generated from benchtop NGS instruments allows for comprehensive sequencing experiments in a single run. HaloPlex can be used to enrich for customized target regions and to prepare libraries for sequencing on Illumina or Ion Torrent in less than 1.5 days with target coverage ranging from a single gene up to the complete exome. The HaloPlex protocol consists of four main steps. In the first step the genomic DNA is fragmented using restriction enzymes followed by the second step where HaloPlex probes are hybridized to targeted fragments. The third step involves DNA ligation of a common primer motif to all targeted fragments which is then used in the final step for multiplex amplification all targeted fragments using one single primer pair. Samples are barcoded during ligation and can thereby be pooled to optimize use of the sequencing capacity on the sequencing chip. To further improve the HaloPlex Exome kit we have developed a new protocol that requires only 50 ng input and that provides coverage of 99% of the target region with less than 5 gigabases of sequencing. Furthermore, we demonstrate a novel protocol where target molecules can be prepared for sequencing without the need for PCR amplification. Using this approach we are able to avoid amplification artifacts by sequencing the actual sample molecules. The content of the exome kit is based on relevant regions from CCDS, Refseq gene, GENCODE, Vega, UCSC, TCGA and MirBase. A total 37 Mb region in targetted in the hybridization reaction which contains over 2 million HaloPlex probes complementary to a redundant set of target regions on the AmpliSeq capture panel. The Access Array System was used to validate results. Results. Targeted capture allowed for decreased bioinformatics processing by focusing only on those gene pathways thought to be pertinent to the disease under investigation. In addition, the pooling design reduced overall sequencing costs. The CRISP algorithm provided a higher sensitivity than BAV in this design. The data suggest a role for targeted NGS of pooled samples when investigating birth defects which are proposed to have a complex and polygenic inheritance.
1643F
Direct selection of microbiome DNA from host DNA. E. Yigit1, G. Feehery1, S. Oyola2, B. Langhors1, L. Apone1, P. Liu1, D. Munafó1, C. Sumner1, J. Bybee1, L. Mazzola2, F. Stewart1, M. Quaill2, T. Davis1, E. Dimalanta1, S. Pradhan1. 1) New England Biolabs, Ipswich, MA., USA; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinton, Cambridge, UK CB10 1SA.

Nucleic acid-based probes such as hybridization, PCR, qPCR and next generation sequencing offer a rapid and highly sensitive option for direct metagenomic detection from specimens when compared with culture-based techniques. Currently, 16S rRNA gene sequencing is the method of choice for metagenomic studies. However, this approach lacks the sensitivity to detect rare members of the microbial community with divergent target sequences, and is not adequate to detect virulence factors in individual strains. Aside from these inherent limitations of amplification and identification of biological samples, the non-microbial host genome itself may interfere with the detection and diagnosis of pathogens due to the higher percentage of host genomic DNA relative to the target microbiome. Therefore, analyses of a metagenome or microbiome directly from host samples by next generation sequencing or PCR are inefficient, difficult and time consuming.

To address this problem, we have developed a unique method for separating large pieces of host DNA from microbial DNA using a methyl-CpG binding domain fused to the Fc portion of a human antibody heavy chain (MBD2a-Fc). This MBD2a-Fc protein is then bound to a protein A magnetic bead and used to separate methylated host DNA from microbial DNA that is unmethylated or methylated at low levels. As a demonstration of the efficacy of this methodology, DNA samples from various sources, such as human saliva, blood, and Plasmodium falciparum contaminated human DNA, were enriched and sequenced on different next generation sequencing platforms. Sequencing data showed that non-microbial host DNA reads decreased 50-fold, corresponding to ~90-95% microbiome DNA in the enriched fraction. Importantly, microbiome diversity after the enrichment remained intact. This simple methodological approach can be used to analyze entire microbiomes in a cost-effective manner utilizing established next generation sequencing platforms, as well as newer single molecule sequencing technologies.

1644W
In solution HLA capture and high-resolution NGS-based typing method and an automated, integrated analysis framework. M. Wittig1, J.A. Anmarkrud2, M. Forster2, E. Ellinghaus1, K. Holm2, L. Wiensbrandt2, S. Sauer2, M. Schmitt2, M. Ziemann1, S. Görg3, T.H. Karlsen4, A. Franke2, 1) Christian Albrechts University, Institute of Clinical Molecular Biology, Kiel, Germany; 2) Norwegian Primary Sclerosing Cholangitis Research Center, Medical Department, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 3) Christian Albrechts University, Institute of Technical Informatics, Kiel, Germany; 4) University of Lübeck, Institute of Transfusion Medicine, Lübeck, Germany; 5) Max Plank Institute for Molecular Genetics, Berlin, Germany.

The human leukocyte antigen (HLA) locus contains the most polymorphic genes in the human genome. These genes play an important role in immune response and much is already known about their role in autoimmunity and infectious disease. The classical characterization of these genes is based on Sanger- or next-generation sequencing (NGS) of a limited amplicon repertoire or labeled oligonucleotides, which identify allele-specific sequences. Using these traditional methods, the rate of possible ambiguities is high and requires manual evaluation of the results, which is also an error-prone process. Here, we introduce a highly automated method, which employs comprehensive insolution targeted capturing of the complete classical class I and classical class II loci in combination with NGS. Our implemented fully automated analysis allowed for the accurate characterization of HLA-A (0.99 allele calling rate), HLA-B (0.99), HLA-C (0.99), HLA-DRB1 (0.98), HLA-DQA1 (0.99), HLA-DQBI (0.99), HLA-DPA1 (0.98) and HLA-DPB1 (0.96). Including possible ambiguities and manual verification allowed for the exact HLA typing of all our reference samples. The reference sample set comprises 261 samples so far and were derived from the International Histocompatibility Working Group (IHWG) biobank and from another German center. The allelic diversity of the reference sample was maximized before enrichment and NGS. For HLA-A we identified 66 different alleles, for HLA-B 106, HLA-C 49, HLA-DRB1 71, HLA-DQA1 20, HLA-DQBI 17, HLA-DPA1 5 and HLA-DPB1 39, respectively. In summary, our method provides a straightforward workflow, which is mainly due to the use of in-solution targeted capturing rather than traditional amplicon-based techniques. The fully automated allele calling delivers high confident allele calls and the number of possible ambiguities is drastically reduced compared to traditional typing (e.g. class I with on average 2.5 possible alternatives per sample). At this very early stage of development, one technician can characterize 182 samples in one week with high confidence and high resolution (6-8 digits).

1645T
The New Sequencer on the Block: Comparison of Life Technology’s Proton Sequencer to an Illumina HiSeq for whole-exome sequencing. J. Boland, M. Yeager, M. Dean, D. Roberson, J. Mitchell, S. Chanock. Cancer Genomics Research Laboratory, NCI Frederick, Gaithersburg, MD.

We assessed the performance of the new Life Technologies Proton sequencer by comparing whole exome sequence (WES) data in a CEPH trio (family 1463) to the Illumina HiSeq instrument. The Proton identified 96% of single nucleotide polymorphisms (SNPs) detected by the HiSeq but only 40% of small insertion and deletion variants (indels). Further comparison of the trio data with Complete Genomics sequence data and Illumina SNP microarray genotypes documented high concordance and accurate SNP genotyping of both Proton and Illumina platforms. However, our study under-scored the problem of accurate detection of indels for both the Proton and HiSeq platforms.

1646F

Virtually all genomic library construction methods use some kind of size-selection in order to avoid adapter artifacts, and to place boundaries on the library fragment size. The gold standard method for size-selection is preparative agarose gel electrophoresis, a procedure that is laborious, irreproducible, and difficult to automate. In response to this need, Sage Science introduced its Pippin Prep automated preparative electrophoresis system in 2010. Since that time, three trends in library construction have been notable: 1) library chemistry has become much more efficient, enabling a reduction in the amount of genomic DNA needed for library construction, 2) the need for high quality genomic libraries in clinical research and medicine has increased dramatically, 3) high quality sequence data are easier to obtain if multiple libraries with different insert sizes are sequenced. To address these trends, Sage is introducing two new preparative electrophoresis systems. The first is a higher-throughput version of the original Pippin Prep system. Like the Pippin Prep system, the high-throughput system is designed to select one or two size fractions from a genomic DNA sample. The channels of the new system have been dramatically reduced in size, a change enabled by the reduced DNA requirements of new efficient library protocols. The disposable agarose gel cassette will fractionate 12 samples per cassette, and the instrument can process two cassettes (24 samples) per run. Run times are typically 0.5-1 hour. The second system is designed to fractionate a single genomic sample into 12 contiguous size fractions. The system uses a two-dimensional process to separate DNA through an agarose column in one direction, and then move the separated DNA fragments sideways into a linear array of elution modules that are positioned alongside the separation column. Each disposable cassette will process a single sample, and the instrument will process two cassettes per run. The instrument can operate in direct current mode for samples up to mid-single kilobases in size, or in pulsed field mode for samples up to 50 kilobases in size. Performance data for the two systems will be described.

The application of Illumina’s sequencing by synthesis (sbs) technology has proved to be very powerful in next generation sequencing and has also been shown to have a lower level of head to head improvements. Illumina has produced platforms that range from systems that: /can generate 600 Gbases of high quality data from libraries in 11 days (HiSeq2000)/produce a 30x coverage of a human genome in less than 27 hours which includes cluster formation on the HiSeq2500 (enable short read (36 bases) data in less than 4 hours or data with paired 250 base reads with the MiSeq system The HiSeq 2500 incorporated features that enable (in the rapid run mode) clusters to be produced on the instrument with the flowcell in situ greatly simplifying the workflow. The instrument can be used as either a high input (>600G) or rapid run (120G in <27 hours) instrument. Recently announced future improvements include extending the read length to 2x250 cycles enabling up to 300G of data to be produced in <60 hours. The MiSeq system has recently been extended to enable paired 250 base sequencing with a high numbers of reads (~7M). It is extremely easy to use, the workflow allowing for cluster formation, paired end and data analysis on instrument and can be combined with a very rapid sample preparation using Nextera technology that takes less than 30 minutes. Run lengths can be selected as needed from single 36 base to paired 250 base reads. The system has also been made easier to use by the incorporation of a Cloud platform called BaseSpace that reduces the need for IT infrastructure. In a further extension of MiSeq the number of runs has been increased to 25M and the run length extended to 2x1200 cycles enabling yields of >15G. Also together with a new chemistry we have been able to demonstrate even longer read lengths. The use of Illumina’s sbs technology is ubiquitous and since its launch in 2006 over 3300 peer cited papers have been published. Here we present some recent advancements in technology and its applications to both whole genome and exome human sequencing.

Towards the 24 hour medical genome. K. Hall1, J. Weir1, S. Humphreys1, Z. Kingsbury1, E. Tsong1, P. Smith1, S. Macarthur1, E.E. Mangulix1, J. Betley1, J. Peden1, N. Miller1, E. Farrow1, L. Willig1, J. Petrikin1, D. Dinwidie2, C. Saunders1, G. Twist1, L. Smith2, S. Soden2, M. Gibson2, S. Kingtonsore1, 1Illumina Cambridge Ltd, Saffron Walden, Essex, United Kingdom; 2Center for Pediatric Genomic Medicine, Children’s Mercy Hospital, Kansas City, MO, 64108.

Clinical settings exist where a point-of-care genome sequence is needed, such as in a neonatal intensive care unit (NICU). 5% of US newborns are admitted to a NICU, and 30% of these may benefit from a rapidly interpreted genome sequence for differential diagnosis of a single gene disorder. To influence medical decision making in NICU babies, a molecular diagnosis must be made very rapidly. For most genetic diseases, a definitive (molecular) diagnosis prefigured for specific treatment, prognostic assessment and genetic counseling. Theoretically, genome sequencing can allow an acutely ill baby to be tested for all 3,736 known genetic diseases at once. Currently, causative associations have been determined between the genetic disease syndromes and 297 regions using the Hiseq-2500 and have previously reported proof-of-concept for 50-hour return of provisional results in 8 newborns. Essential to this is electronic entry of clinical features by the ordering physician, thereby defining a superset of differential diagnoses and genomic regions to be tested. We report that we have now expanded this clinical genome ordering interface to structured vocabularies of over 5,000 diseases and 8,000 clinical terms. We also report improvements in the overall workflow that shorten return of provisional test results to 24 hours. In particular, methods have been developed that shorten the times for sample preparation, sequencing, alignment and variant calling. We will also demonstrate the use of 2 x 250 read lengths for better alignment in repetitive regions and identification of polynucleotide mutations. Finally, we report clinical utility in 25 cases.


Droplet Digital PCR reduces biases and improves representation of amplicons in next generation sequencing libraries. The amplicons generated by multiplexing assays are improved when partitioned, compared with standard single tube NGS and MiSeq NGS methods. Partitioning the sample into droplets reduces biases that arise in PCR such as competition between assays. Custom multiplexed assays as well as standard commercial cancer panels were tested for improvements in read coverage when comparing standard workflows with Droplet Digital PCR. Here we present a facile methodology which easily integrates in to current NGS amplicon library workflows for improvement in reducing amplification bias.
1652F

Comparison of enzymes, shear time and capture products to improve whole exome sequencing workflow. B. Marosy, B. Craig, K. Hetrick, H. Ling, A. Robinson, S. Griffith, J. Romm, K.F. Dohnen. Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR is continually seeking new ways to improve its workflow and generate high quality data. A major obstacle in next generation sequencing is the ability to amplify through GC/AT rich regions of the genome. Current methodologies for library preparation and targeted capture rely on PCR to enrich for the prepared libraries. In addition, increased sequencing read length capability has also required changes to the workflow to increase insert sizes. Here we tested two different enzymes (Agilent© Hercule II Fusion DNA Polymerase & Kapa Biosystems HiFi DNA Polymerase) and implemented a different shearing strategy to optimize our current workflow. We also applied these improvements to a Nimblegen TM workflow to compare exome captures between the Nimblegen SeqEZZCpTM v3 and the Agilent SureSelectTM XT Human All Exon v4. Four unrelated HapMap samples with DNA inputs of 500ng were processed in parallel using 1) Agilent SureSelect XT library prep (XT) & Hercule enzyme; 2) XT library prep & HiFi enzyme; 3) XT library prep & HiFi Enzyme w/shearing modifications; 4) Kapa Library prep & HiFi enzyme w/shearing modifications. Library methods 1, 2 and 3 were hybridized for 24hrs using the Agilent v4 exome and library method 4 was hybridized for 72hrs using the Nimblegen v3 exome. All libraries were clustered using the Illumina® cBOTTM Cluster Generation system. One hundred bp paired end sequencing was performed on the Illumina HiSeqTM2000 platform. Sequencing data was downsampled to ensure an equal comparison between methods. In addition, a product-neutral UCSC exon bed file was used as the ‘on target’ bed file to more accurately compare the capture products. Preliminary sequencing data analysis of methods 1 and 2 yielded 95.8% and 95.5% of 20 bp reads of targeted bases covered >=8x; a molecular duplication rate of 5.1% and 4.8%; mean insert size of 175bp and 193bp; respectively. GC plots indicated improvement when using the HiFi enzyme for normalized coverage and base quality at GC/AT rich regions. We propose that these improvements will enable researchers to more consistently capture regions of the genome.

1653W

Simplified and improved methods for preparing high quality genomic libraries for use on Illumina® sequencing systems. V.P. Smith, S.J. Humphray, R.M. Sanches-Kuiper. Illumina UK Ltd, Nr Saffron Walden, Essex, United Kingdom.

Simple, affordable and reliable methods for the preparation of high quality genomic libraries are essential for cost-effective whole genome sequencing (WGS). When selecting a library preparation method for WGS, key performance metrics to consider include (i) simplicity and duration of the workflow; (ii) DNA input requirement; (iii) yield of the final library; (iv) uniformity of coverage and (v) library diversity or complexity. We will present results obtained with the latest versions of Illumina’s TruSeq® DNA Sample Prep workflows. TruSeq PCR-free and TruSeq Nano, used in combination with the HiSeq®2500 and MiSeq® sequencing systems. The PCR-free method produces robust yields from moderate DNA input quantities and generates very high quality genomic libraries with sufficient diversity to support deep sequencing of human and other large genomes. Elimination of PCR-induced biases results in libraries that exhibit uniform coverage across the full range of GC contents . The TruSeq Nano workflow produces relatively unbiased, high yield and high diversity libraries from large and complex genomes starting with as little as 100ng of DNA. These workflows are gel-free, straightforward to execute, automation-compatible and can easily be completed in less than a day, making them ideally suited to the fast turnaround times of the HiSeq 2500 and MiSeq systems. The difficulty of meeting the performance requirements described above is increased substantially when working with degraded DNA such as that extracted from FFPE samples. We will also report on the development of methods for obtaining high quality WGS data-sets from practical quantities of FFPE DNA.

1654T

Successful whole-exome sequencing of genomic DNA isolated from preserved mixed-placental tissues. M.K. Veerapen1, L. Pelaez2, M.M. Rodriguez3, J.E. Potter4, E. Rampersaud1, O.A. Bodamer1, 2, 5 Hussman Institute of Human Genetics, University of Miami, Miami, FL; 2) Department of Biochemistry and Molecular Biology, University of Miami, Miami, FL; 3) Department of Pathology, University of Miami, Miami, FL; 4) Department of Obstetrics and Gynecology, University of Miami, Miami, FL; 5) Department of Biostatistics, University of Miami, Miami, FL;

Studies on the genetics of obstetrics and gynaecology have previously focused on utilizing fresh placental tissue (PT) for downstream genetic analysis. However, fresh PT would rarely be histopathologically evaluated for deep-phenotyping purposes and cannot be stored for long-term. Therefore, the use of fixed and paraffin embedded PT would be of interest due to the increasing importance for retrospective genetic studies. With the technological advancement of DNA isolation methods and next-generation sequencing, this could increase the efficiency in variant identification from the preserved PT samples. Due to the fact that the placenta contains a mix of maternal and fetal tissue, the experimental and variant identification challenges are significantly increased. We propose to isolate DNA from preserved PT and successfully perform variant identification. We obtained deidentified PT samples at non-crosslinking UMFix fixed and paraffin embedded (UMFFPE); foetal membrane (FM) and umbilical cord (UC) samples of mother-foetal (son) dyad pairs from Jackson Memorial Hospital, Miami, FL. The UMFFPE-UC 20 μm curls, containing only foetal surface, were processed for DNA isolation. The UMFFPE-FM containing both maternal-foetal (son) surfaces were separated using laser-capture microdissection (LCM) and DNA isolation. Genomic DNA (gDNA) obtained from UMFFPE-LCM-FM and UMFFPE-UC were then processed for library construction, capture and hybridization using the Roche Nimblegen V3 and Agilent SureSelect 50 Mb V4 respectively for sequencing. Paired-end reads were mapped and aligned with the Burrows-Wheeler aligner according to the Genome Analyzer Tool Kit (GATK) v2.5 pipeline. We successfully isolated gDNA from both UMFFPE-LCM- FM and UMFFPE-UC samples of good gDNA yields and quality. The UMFFPE-LCM-FM samples tested negative for foetal contamination using a PCR-based method. Preliminary sequencing results have shown a capture of an average of 78%; and a read of depth of 71%; at 20X. To the best of our knowledge, this is the first report of the successful whole-exome sequencing and gDNA isolation from preserved PT samples. Potentially, this protocol can be utilized for retrospective genetic studies related to obstetric conditions such as preterm births, intrauterine growth restrictions, intrauterine foetal demise, preeclampsia and choioamniotitis.

1655T


Targeted resequencing provides a cost-effective approach to identify variants by isolating specific genomic regions of interest within a whole genome library. However, high-throughput sample processing and fast time-to-answer can be affected by lengthy library preparation and enrichment procedures. Here we describe the Nextera Rapid Capture Enrichment assay that allows researchers to go from genomic DNA (gDNA) to targeted enrichment data in less than 2.5 days when accompanied by sequencing on a HiSeq2500. By requiring only 18 hours of hybridization time, a single enrichment wash solution, and rapid fluorometric sample quantification procedures, sequencing ready samples can be prepared in 1.5 days with no more than 5 hours of hands-on time. We present data, from a familial CEPH pedigree trio (Mother NA12892, Father NA12891, Child NA12878), showing how the Nextera Rapid Capture Enrichment assay can be used to target ~37 Mb of focused exonic content that covers at least 98% of coding regions. Pooled Oligo pool library (SureSelect XT & Herculase enzyme; 2) XT library prep & HIFI enzyme; 3) XT library prep & HIFI Enzyme w/shearing modifications; 4) Kapa Library prep & HIFI enzyme w/shearing modifications. Library methods 1, 2 and 3 were hybridized for 24hrs using the Agilent v4 exome and library method 4 was hybridized for 72hrs using the Nimblegen v3 exome. All libraries were clustered using the Illumina® cBOTTM Cluster Generation system. One hundred bp paired end sequencing was performed on the Illumina HiSeqTM2000 platform. Sequencing data was downsampled to ensure an equal comparison between methods. In addition, a product-neutral UCSC exon bed file was used as the ‘on target’ bed file to more accurately compare the capture products. Preliminary sequencing data analysis of methods 1 and 2 yielded 95.8% and 95.5% of 20 bp reads of targeted bases covered >=8x; a molecular duplication rate of 5.1% and 4.8%; mean insert size of 175bp and 193bp; respectively. GC plots indicated improvement when using the HIFI enzyme for normalized coverage and base quality at GC/AT rich regions. We propose that these improvements will enable researchers to more consistently capture regions of the genome.

Posters: Bioinformatics and Genomic Technology
Comparison of conventional and PCR-free library preparation methods for next generation sequencing. P.D. Witmer, B. Marosy, B. Craig, K. Hetrick, A. Robinson, K. Doheny. Center for Inherited Disease Research (CIDR), Johns Hopkins University School of Medicine, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics con-
sultation to investigators working to discover genes that contribute to human disease. As part of our continuing effort to implement emerging NGS technol-
ologies, we now offer whole genome sequencing (WGS) as a complement to whole exome sequencing (WES) within the suite of services available at CIDR. Furthering our aim to evaluate improvements in methodology, we have been focusing on DNA library preparation methods for sequencing on the Illumina HiSeq2500 platform. Standard library preparation methods include a PCR enrichment step prior to cluster generation. Biases inherent in PCR amplification result in uneven read coverage and increase the num-
bbers of duplicate fragments present in the library. For the HiSeq instruments in particular, coverage of sequence reads is known to be notably lower in GC rich regions, which can affect mapping quality and variant calling. To compare library preparation methods that employ PCR directly to methods that do not, we constructed libraries for NGS with DNA from a well character-
ized HapMap Trio (NA12878, NA12891 & NA12892) using a conventional DNA library preparation kit from Kapa Biosystems and the TruSeq DNA PCR-Free Sample Preparation Kit available from Illumina. Our QC metrics indicate that library size increased using the PCR-free kit (2.5 fold), improving the complexity of the sample. We also observed significant decreases in read pair duplicates (2 fold) and chimeric molecules (3 fold) compared with samples enriched by amplification. GC plots demonstrate more uniform coverage for PCR-free libraries albeit with a slight reduction in the total number of bases covered above 20X. For mean coverage, TuTv, homozy-
gous concordance and sensitivity to SNP genotyping arrays, there were no appreciable differences between the two methods. However, for PCR-free libraries, there were notable increases in unmapped reads (3 fold) and number of duplicate reads (2 fold) and the percentage of reads containing adapter sequences, which likely reflect broken mate pairs and a smaller insert size (350 vs. 460 bp). Although the omission of PCR shortens the workflow for library generation, this protocol has limitations. This report highlights this approach as an effective tool.

Accurate Modeling of Indel Genotype Likelihoods from Sequencing Data. A. Tan, HM. Kang, GR. Abecasis. University of Michigan, Ann Arbor, MI.

Short indels are the second most common class of genetic variants. Next generation sequence data and alignment-based indel calling algorithms allow efficient discovery and genotyping across a large number of samples, while retaining high power to detect rare variants. However, the current state-of-the-art methods for analysis of indels are not as accurate as their counterparts for analysis of SNPs. One of the reasons is that indels vary much in length and sequence complexity and thus require more careful handling to cater for the variety of forms. Here we address the problem of bias towards reference alleles when modeling the likelihood of short indels. We observed that in widely used indel calling algorithms, genotype likelihoods lean preferentially towards reference alleles, and the bias becomes larger as allele length increases for indels. For example, for high quality 5-
bp indel sites, the proportion of reference alleles in aligned reads from individuals with heterozygous genotypes ranges from 53% to 73%, com-
pared to the expected value of 50%. In our improved algorithm, we carefully construct appropriate local haplotypes with respect to each candidate allele and use a local hidden Markov model that keeps track of flanking sequences and insertion/deletion fragments to qualify and quantify the evidence for each allele. Our procedure allows us to attain an allele balance close to 50% and reduce genotype likelihood discordance substantially compared to analyses using GATK, samtools, and other callers; particularly for longer alleles. As a result, our model also provides us with improved filtering of variants as it results in a more robust distribution of variant features against the confounding due to allele length. Our indel pipeline performs variant calling, genotyping, and filtering while modeling contamination and systematic alignment artifacts. The complete process is incorporated in the updated version of GotCloud pipeline. Along with other independent efforts to improve the existing indel calling algorithms, we anticipate that our method will con-
tribute to increasing our understanding of the genetic risk factors beyond single nucleotide variants.

A complete workflow from sample preparation to analysis using Sure-
Select target enrichment system for Ion Proton semiconductor sequencing. K. Jeong1, J. Ong2, E. Agre2, F. Karlsson2, A. Ashutosh2, C. Cocc3, F. Useche1, J. Ghosh1, H. Johansson3, S. Hape2, D. Roberts1, 1 Agilent Technologies, Santa Clara, CA; 2 Agilent Technologies, Cedar Creek, TX; 3 Agilent Technologies, Upplands Väsby, Sweden.

Whole exome or targeted sequencing for protein-coding regions has pro-
vided a cost effective way to identify common and rare polymorphisms that are associated with Mendelian disorders and complex diseases. With increased capacity of semiconductor sequencing, highly multiplexed sam-
iples can be studied in a single sequencing run. However, a complete work-
flow processing raw DNA samples to identify DNA variants in target regions is not easily accessible. Here we describe an analysis workflow to study multiplexed samples in semiconductor sequencers for several target sizes: 50Mb (Human All Exon), 3.2Mb (Human Kinome) and a 1Mb custom design. The workflow includes library preparation, SureSelect target enrichment, semiconductor sequencing, and variant calling with SureCall software (beta version). Improved and simplified steps for library preparation and target enrichment maximize multiplexing and produce consistent results in the Ion Proton sequencer. Sequencing output can be visualized and summarized in a report with SureCall which is optimized for use with Agilent’s target enrichment system. We demonstrate high capture efficiency, uniformity, and reproducibility of enrichment. The results from different capture sizes show comparable performance regardless of various targeted designs. The combination of efficient target enrichment system, semiconductor sequencing, and SureCall software provides a fast and convenient tool to assess DNA variants in genomic regions of interest.


Gene regulation is central to the life of a cell. Sophisticated programs of gene expression trigger developmental pathways in response to environ-
mental stimuli and cell disorders. Even in nominally homogeneous cell popu-
lations, cell-to-cell differences are observed. These differences represent different gene expression patterns driven by modular gene regulation. How-
ever, due to technical limitations, biologists traditionally investigate gene
regulation by utilizing bulk-scale methods to measure average gene expres-
sion values in populations of cells. Therefore, the ensemble gene expression profile of a population may mask the behavior of gene regulation in any individual cell. The Fluidigm C1™ Single-Cell Auto Prep System allows researchers to investigate whole transcriptome profiling of individual cells with a simple, automated workflow. In this study, we used the C1 System to prepare mRNA-Seq libraries from >10 different cell types (e.g. HL60, iPSC, heart primary neuronal cells), with an average sample size of 24 single cells per cell type, and we then conducted deep sequencing on the Illumina HiSeq platform. Sequence depth of 5-10 million reads per cell was obtained and systematically analyzed for gene expression profile correlation at the single-cell level across diverse cell types. We found that extremely deep sequencing (10-30 million reads) was not required to conduct differential gene expression analysis, and that even with read depths of only 1 to 5 million reads, we were able to develop robust gene expression profiles from single cells. Our results indicate that many highly correlated gene expression profiles can only be observed at the single-cell level and are masked in bulk data. Furthermore, the data revealed that gene correlations in single cells are cell-type dependent while highly correlated genes in single cells enrich for relevant biological pathways. Using the validated transcription factor binding site (TFBS) data from the ENCODE project, we further confirmed that some of these highly correlated genes share common TFBS in their promoter regions. Our study suggests gene co-profiling analysis at single-
cell resolution is a powerful tool that can help biologists precisely dissect gene regulation networks and better understand cellular heterogeneity.

Posters: Bioinformatics and Genomic Technology
High throughput DNA sequencing allows whole human genomes to be resequenced rapidly and inexpensively producing a comprehensive list of variants relative to the reference genome. However, short read sequencing technologies are limited in their ability to determine phasing information, thus resulting in heterozygous calls being represented as the average of the maternal and paternal chromosomes. Phasing information is of critical importance to personal medicine as it provides a better linkage between genotype and phenotype, permitting new advances in our understanding of compound heterozygote linked diseases, pharmacogenomics, HLA typing, and prenatal genome sequencing. Here, we describe a new sample prep method that enables whole human genome haplotyping at high accuracy using only 300Gb of sequence data. Genomic DNA was fragmented into ~10Kb fragments, end repaired, and ligated to adapters. Hundreds of aliquots with approximately 50Mb of DNA in each were amplified, fragmented and converted into individual shotgun libraries. The pooled libraries were sequenced in a single lane of a HiSeq2500 at 2×100bp to generate ~30Gb of sequence. The resulting sequence information was analyzed to obtain a set of long blocks of ~10Kb, covering multiple heterozygous SNPs, allowing phasing of these SNPs relative to each other. An HMM-based phasing algorithm was used to compute the most likely phase and confidence intervals based on the observed coverage and sequence quality scores. Phasing of these blocks relative to each other was done by another HMM-based algorithm which uses a panel of previously phased genomes. Comparing our results with phase information derived by transmission from the parents, we found that over 98% of heterozygous SNPs were phased within long blocks (N50=500kb) at a switch error rate below 1 switch per megabase of phased sequence. We present results obtained from multiple cell lines and human samples. This new library prep method and data analysis pipeline enables whole human genome phasing with only 30Gb of raw sequence, which represents only ~30% more sequencing than current 30x baseline coverage thresholds by as much as 20%, with median per-base coverage rising from 88.4% to over 99.9%. Furthermore, comparison to coverage levels reported on the Exome Variant Server showed comparable or, in some cases, superior coverage across targeted regions. Independent of MIP design, a series of optimizations of the capture protocol improved the relative efficiency of capture of high GC targets tenfold or greater. This work demonstrates that improved modeling of the features affecting MIP capture can be used to successfully predict relative MIP capture efficiencies and, coupled with optimizations of the capture protocol, have the potential to push coverage of target sequences close to 100%. Further improvements through variance modeling enable the application of MIP-based sequencing to clinical diagnostics.

Next Generation Sequencing (NGS) has significantly impacted human genetics, enabling a comprehensive characterization of the human genome as well as a better understanding of many genomic abnormalities. By delivering massive DNA sequences at unprecedented speed and cost, NGS promises to make personalized medicine widespread in the foreseeable future. To date, laboratory workflow integration with clinical sample preparation remains primarily due to the limited quantities of sample DNA available, but also due to the low quality of some samples, such as FFPE samples. To overcome this challenge, we have developed NEBNext® Ultra DNA Library Prep Kit for Illumina® for Illumina®; a fast library preparation method using novel reagents and adaptors, including a DNA polymerase that has been optimized to minimize GC bias. This method enables library construction starting with nanogram quantities of DNA, and can be used for both intact and fragmented DNA, such as that found in FFPE samples. Moreover, the workflow is compatible with multiple NGS platforms, enabling use of a single library prep kit for creation of a variety of sequencing libraries. Illumina, 454, and SOLiD 5500 libraries were successfully generated and sequenced on the corresponding platforms to produce high quality sequencing data.


In order to generate a broad and structurally diverse adaptive immune repertoire, T and B-cell receptor genomic loci undergo non-precise somatic V(D)J segment rearrangements, including templated and non-templated nucleotide additions and deletions. Several studies observed that even in healthy subjects certain TCR V segments are commonly utilized while others are quite rare, also thereby reflecting a snapshot of the currently active adaptive immune system. A shift towards mono- or oligoclonoity and the associated usage of dominant TCR V segments is often observed in infection, chronic inflammation, autoimmune disease and cancer. Recently, Zhang et al. described a Nanostring nCounter based "direct TCR expression assay" that allows a rapid and sensitive screen for the usage frequencies of all functional TCR Vα and Vβ segments. Here we report on the modification of the initially described codset for inclusion of TCR Vγ and Vδ segments as well as the exemplary application for TCR Vγ and Vδ segments screening in colon and skin biopsies obtained from Crohn’s disease and psoriasis patients. As TCR transcripts make up only a minority in total RNA from biopsies, our analysis involved PCR amplification of TCR transcripts and re-transcription to cDNA for subsequent Nanostring nCounter analysis. We compare the obtained patterns of TCR V segment usage to corresponding profiles generated by massive parallel immune repertoire sequencing. In summary, profiling of TCR V segment usage in tissue biopsies using Nanostring nCounter enables the rapid and efficient detection of expanded T cell clones associated to disease.

Evaluation of whole genome amplified DNA and reduced genomic DNA for high performance of Illumina SNP microarrays. C.L. Dagnall, H.C. Morton, D.D. Hicks, W. Zhou, D. Deng, M. Yeager, S. J. Chanock. 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD; 2) Cancer Genetics Research Laboratory, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD.

The high genomic DNA input requirements for whole genome single nucleotide polymorphism (SNP) microarrays can limit the scope of molecular epidemiological studies. We evaluated alternatives to reduce input DNA requirements for Illumina Human Infinium® assays. Whole genomic DNA (gDNA) amounts and whole genome amplified DNA (wGA DNA) as inputs into the Infinium® assay protocol. These alternate inputs reduce gDNA template requirements by 75% and 96.67%, respectively. 192 DNA samples from 71 individuals were obtained from multiple biopsiemen sources and genotyped using standard and alternative inputs. When compared to genotypes obtained from the standard input amount, we observed 99.98% and 99.73% median concordance with reduced gDNA and wGA DNA, respectively. We demonstrate that carefully conducted studies with alterative inputs can yield high-quality genotyping results. These findings enable investigators to consider expansion of ongoing studies, including genome-wide association studies, challenged by small amounts of available DNA.


BRAF V600E mutations are present in a high percentage of papillary thyroid carcinomas and malignant melanomas. Problems in routine diagnos- sis include limited sample quantities, ambiguous results, and the presence of PCR inhibitors, particularly in melanomas. A sensitive, rapid, and inexpensive method for detecting these and other mutations that is not affected by these limitations would be of great utility. Several papillary thyroid carcinoma and malignant melanomas from formalin fixed and native samples were analyzed using both a highly specific amplification refractory mutation system (ARMS) and digital PCR. Comparison of results indicates that digital PCR on the QuantStudio™3D System is able to detect BRAF V600E mutations in all samples where the mutations were detected by ARMS PCR, even when preamplification was required for mutation detection in some samples. Digital PCR was able to detect mutations in samples from melano- mas that contained PCR inhibitors. The digital PCR results are also quantita- tive, allowing the determination of the mutant to non-mutant ratio of the sam- ples.


Several major challenges have impeded the widespread application of targeted sequencing assays in the clinical setting. Foremost among these challenges are the high cost and complexity of sample preparation for the next generation sequencing platforms. Current technologies for target enrichment require between two to four days of preparation before samples are ready for the sequencing run. Furthermore, in order for current solutions to approach cost effectiveness, many samples need to be batched in a single run, adding to the complexity, increased likelihood of error and increased turnaround time. The GnuBIO unique desktop sequencer uses a single-use cartridge-based consumable that incorporates all of the preparation steps that are typically performed separately including target enrichment, amplifi- cation and sequencing. This drastically simplifies clinical workflows and reduces the operator hands-on time to less than 5 minutes, an unprece- dented total. Biochemical reactions take place in a cascading manner, with individual reactions contained inside minute emulsion droplets, which flow through microfluidic channels where they are injected with assay reagents, incubated at the required temperatures, and finally detected for sequencing. Each droplet results in a single DNA sequence read, with reads up to 1000nt long. GnuBIO is collaborating with the City of Hope to develop an oncology pathology gene panel for detection of actionable mutations on the GnuBIO platform. The system is designed to achieve the following: /Starting material: genomic DNA from FFPE samples or other types of samples. /Total run time of less than 4 hours /Sample preparation requires only genomic DNA extraction /DNA library quantitation and sequencing. /Robust detection of low frequency alterations (~1%-5%). /High accuracy with average per base accuracy of 99.999% /Ease of integration with existing analytic and sample tracking workflows. /Low per sample cost - $200 per sample /Low instrument cost - $50,000. In addition, the cartridge is a closed system, and can be readily adapted to new panels. The cartridge is a dry consumable, containing all reagents needed to run a single DNA sample, reducing the risk of contamination and error.


Technological advances have significantly increased the sensitivity of genomic assay systems. Such advances have placed a premium on lab procedures that keep samples contaminant free. That requirement for sample transfer is usually satisfied with costly single use plastic tips. As an efficient and cost effective alternative, IonField Systems has developed the TipCharger™ plasma treatment system, a microplate sized module that uses cold plasma to clean plastic nitrotype tips. Data shows that a 30-second plasma treatment removes DNA to undetectable levels without generation of any liquid or solid wastes. Use of the TipCharger can provide a valuable altera- tion that will speed analysis and decrease costs associated with modern genomic analysis.
1671W

Next generation sequencing (NGS) enables analysis of any or all parts of the human genome in a way that opens new opportunities for clinical research and testing that is at more comprehensive, faster and less expensive than previously possible. For diagnostic applications that require deep sequencing such as cancer mutation discovery or detection from FFPE tissue or plasma, target enrichment by amplification or capture is critical to reach the goals for sequence accuracy and economy of sequencing time and reagents. Target capture using Agilent SureSelect or Roche NimbleGen SeqCap EZ reagents use a recommended amount of about 1000 ng of unamplified input DNA for a successful outcome. Thus it is difficult to use those methods on unamplified FFPE or plasma samples, because the amount of total DNA available is only 1 - 300 nanograms of total DNA. Manufacturer-recommended library amplification protocols specify at least 100 ng input. We have optimized library synthesis, amplification and capture protocols to achieve excellent whole exome and whole kinome results using less than 10 ng unfixed tissue or plasma DNA inputs, or larger quantities of FFPE DNA. We used the Rubicon ThruPLEX-FD library reagents, the Agilent and Roche target capture reagents, and modified protocols (when necessary). Sequencing metrics such as percentage reads on target, target coverage, and SNP or variant calling were equivalent to results using unamplified libraries or other library kits using substantially larger amounts of input DNA recommended using manufacturer-recommended protocols. Libraries made from the smallest amounts of input did produce the expected higher percentages of duplicate reads. The optimized reagents and protocols will be available to researchers and will enable high-quality targeted sequencing to be achieved starting with substantially less input of clinical DNA samples.

1672T
SureSelect strand specific RNA library prep kit provides a fast and streamlined workflow for preparing directional libraries from total RNA. B. Arezi1, B. Huse2, F. Useche2, A. Tsalenko2, B. Novak2, A. Lucas2, K. Chen1, H. Tang1, H. Hogrefe1. 1) Agilent Technologies, La Jolla, CA 92037; 2) Agilent Technologies, Santa Clara, CA 95051.

RNA-Seq is a revolutionary technology for whole transcriptome analysis. Deep sequencing of cDNA has been used to quantify transcript levels, confirm gene annotation, and identify novel transcripts, splice variants, and SNPs. RNA-Seq protocols that preserve strand information are critical for identifying antisense transcripts that play a role in gene regulation, determining the exact boundaries of genes transcribed on opposite strands, and accurately measuring expression levels of overlapping genes. Here, we present the SureSelect Strand Specific RNA Library Prep Kit, which includes all reagents required for isolating polyA RNA and preparing directional RNA libraries for Illumina sequencing from as low as 50ng total RNA, using the dUTP marking method (Parkhomchuk et al (09) NAR 37: e123). In a comprehensive comparison of directional RNA-Seq methods, the dUTP marking method was found to provide the most compelling overall balance across all sequencing metrics examined, including strand specificity, library complexity, unevenness and continuity of coverage, and accuracy of gene expression profiling (Levin et al (10) Nature Methods 7: 709-715). With our streamlined automation-friendly protocol, which includes master mixes and combined enzymatic steps, researchers can prepare directional cDNA libraries with >99% strand specificity from total RNA in about 5 hours, without the need for gel size selection. A comparison of RNA-Seq to gene expression array shows good correlation (R>0.8) between differential gene expression ratios (MAQCA and B) obtained using Agilent’s SureSelect Strand Specific RNA Library Prep kit and SurePrint G3 Human Gene Expression Microarray 8x60K. Finally, we show that SureSelect Strand Specific RNA Library Prep is compatible with target enrichment by performing capture using the SureSelect Human Kinome bait library. Enriched libraries show high target and strand specificity (80% on-target reads; >99% strand specificity), and correlation of mean gene RPKM values between target-enriched and un-enriched libraries (R=0.8).

1673F

Once thought homogeneous populations of cells have recently revealed a startling amount of gene expression variation. New understanding has largely been driven by recent advances in the sensitivity of technologies such as next generation sequencing and sequence detection analysis, which are now capable of analysis at the single cell level. Here we describe how the power of flow cytometry can be exploited to isolate, characterize, phenotype, and purify multiple cell populations at the single cell level. We describe a complete workflow starting with heterogeneous cells that can be isolated at rapid speed by single cell depositions into 96-well plates using the newly developed BD FACSAria™ flow cytometer from BD Biosciences. We demonstrate the use of new analytical tools such as index cell sorting, which can help track those individual cells during and after sorting. Finally we describe how, using single cell gene expression with the NanoString® platform, we can demonstrate high fidelity of the sorted cells.

1674W

The Ion Torrent PGM has made it possible to generate large amounts of sequence data in relatively short run times. Consequently, it has become increasingly important to be able to generate large numbers of libraries from a variety of samples quickly and inexpensively. Here we present a fast workflow for construction of RNA libraries for whole transcriptome analysis. miRNA libraries can accommodate low total RNA input with no need for enrichment and are free of adapter-dimer contamination. miRNA libraries retain strand specificity for accurate gene expression quantification and discovery of antisense transcripts. The strand-specific libraries are compatible with poly(A) mRNA as well as ribosomal RNA-depleted total RNA. Using this simple and rapid procedure, library construction time is reduced to one day and the yield of the library is significantly improved. To reduce cost and increase sample throughput, libraries can be barcoded during amplification. The multiplexed libraries can then be pooled before size selection, reducing the number of steps in the workflow.

1675T

Identification and analysis of small RNA by deep sequencing requires preparation of a di-tagged cDNA library. Most library preparation methods for di-tagged cDNA are based on sequential ligation of adaptors. The excess of adaptors required to provide sufficient library yield leads to adaptor-dimer formation which strongly contaminates the library (unless several gel purification steps are performed to remove unligated adaptors). We have developed a novel method to generate di-tagged small RNA libraries free of adapter-dimer contamination without introducing any additional enzymatic steps or gel purifications. Using this simple and rapid procedure, library construction time is reduced to one day and the yield of the library is significantly improved by enabling the addition of a high concentration of adaptors, thereby increasing the percentage of small RNA molecules included in the library. To reduce sequencing cost and increase sample throughput we have developed a barcode strategy to tag samples during library construction. Up to 24 multiplexed libraries can then be pooled together before size selection, reducing the number of steps in the workflow. Additionally, gel-free size selection greatly simplifies the overall workflow. Our method for generating multiplexed small RNA libraries reduces bias by ligation, increases representation of modified small RNAs and simplifies the workflow during library construction for small RNA analysis and discovery.
Molecular Indexing for Improved RNA-Seq Analysis. M. Toluie, J. Ris-inger, P. Nakasthe. BioX Scientific, Austin, TX.

Most current Next Generation Sequencing (NGS) library prep methods introduce significant sequence bias. The use of enzyme processing and fragmentation steps can introduce errors in the form of incorrect sequence and misrepresented copy number. Conventional RNA sequencing library construction involves the ligation of a population of cDNA molecules with adapters prior to amplification and sequencing. An inherent weakness of conventional RNA-Seq analysis is that cDNA fragments that amplify more efficiently will unavoidably result in a higher number of reads than cDNAs that do not amplify as well during the library construction PCR step. Therefore, when multiple reads mapping to the same transcript are encountered, it is not possible to determine whether sequenced reads originate from the same or different cDNA molecules. With molecular indexed libraries, each molecule is tagged with a molecular index randomly chosen from ~10,000 combinations so that any two identical molecules become distinguishable (with odds of 10,000/1), and can be independently evaluated in later data analysis. Analysis using molecular indexing information provides an absolute, digital measurement of gene expression levels, irrespective of common amplification distortions observed in many RNA-Seq experiments. This type of indexing requires no additional steps in RNA-Seq workflow and increases the precision of downstream analysis. At low sequencing depths, analysis use of molecular indices is identical to conventional analysis and generates equivalent RPKM values in all applications. As sequencing depth increases, individual molecular resolution also increases. In quantitative RNA-Seq experiments, the molecular indices distinguish re-sampling of the same molecule from sampling of a different molecule. At high sequencing depths, each molecule can be distinguished and the entire library can be analyzed to provide absolute numbers of each molecule. Resolving individual clones of molecules is critical for increasing sequencing accuracy, measuring bias, PCR duplication rates and identifying mutations in complex sample types. While it is well known that library prep methods introduce bias, tools for measuring it are needed if we are to start using NGS for accurate and quantitative gene expression measurements. Toward achieving that goal, we propose the use of molecular indices for all RNA-Seq experiments.


The lack of standardization and high quality samples from biopositories can impede the progress of disease and basic research exposing a need for quality control and assessment methods to determine sample utility. The Fluidigm biobanking panel facilitates quality control and assurance studies on DNA samples. The panel consists of 96 assays selected by Dr. Andrew Brooks of Rutgers University to provide critical information regarding sample identity, integrity and quality. In this set of experiments, we will determine the limit of detection for gender contamination, assess DNA degradation detection and evaluate the performance of low quality samples assayed by the Fluidigm biobanking panel on the 96.96 genotyping integrated fluidic circuit in the BioMark™ HD system.


The use of high-density DNA microarrays for accurate, cost-effective genotyping of single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms (indels) plays an important role in the identification of the underlying genetic basis of common, complex human diseases. As the catalog of both common and rare variants in multiple worldwide populations continues to expand, microarrays are well positioned to capitalize on this evolving information content for targeted genotyping. Here we present an overview of Axiom® 384HT Genotyping Solution in which 384 individual arrays, contained in the footprint of a standard microtiter plate, offer the capability to simultaneously genotype 384 samples at 50,000 variants per sample with a throughput greater than 3,000 samples per week. The solution thus offers a novel technological capability to genotype large numbers of markers at a very high sample throughput in a cost-effective manner. Axiom® Genotyping Solution enables complete automation of DNA target preparation, including DNA amplification and enzymatic fragmentation of post-amplification products, on liquid-handling workstations. Following target preparation, arrays are processed using GeneTitan® Multi-Channel (MC) Instrument. The Axiom® myDesign™ Custom Arrays in the 384-layout can be designed with markers from the Axiom® Genomic Database or from de novo SNP discovery initiatives. Axiom 384HT Genotyping Solution retains full compatibility with the existing Axiom instrumentation platform and downstream data analysis software. Genotyping performance consistently achieves an average sample call rate ≥99.0%, average sample concordance to independent DNA genotype information (HapMap) ≥98.9%, and intra- and inter-run reproducibility ≥98.9%. In summary, new Axiom advances with the 384-array layout further extend the platform’s capabilities for highly multiplexed genotyping of variants in a single assay. This offers high sample throughput coupled with minimal manual intervention enabling multiple applications including biomarker discovery, sample QC and tracking for large cohorts, and post-GWAS fine-mapping and causal variant analysis.

Successful Illumina Infinium Beadchip high-density genotyping from fragmented and low concentration samples. J.C. Tackney1, D.J. Witherspoon2, L.B. Jorde2. 1) Dept. of Anthropology, University of Utah, Salt Lake City, UT; 2) Dept. of Human Genetics, University of Utah, Salt Lake City, UT.

Currently, high-density and high-quality SNP genotype data cannot be consistently assayed from small amounts of degraded DNA samples. DNA quality and concentration remain a bottleneck both in SNP microarray and next-generation sequencing library creation. While sequencing libraries have successfully been processed from degraded ancient DNA samples, costs are still prohibitively expensive for wide use in forensic DNA laboratories. Generating SNP genotype data from such samples remains a goal in identification of human remains and in crime scene investigations.

Towards that end we created panels of human DNA samples varying in concentration and fragmentation profiles. We first processed these samples through Illumina’s proprietary Infinium FFPE DNA Restoration Solution and genotyped them on the HumanOmnierpress-FFPE BeadChip. We determined that we could genotype 1 ng of DNA fragmented to 300 bp with an ~84% call rate and, when compared to genotypes from 100 ng of non-fragmented control DNA, a ~2.2% discordant call rate. Obtaining more than 500,000 usable SNP genotypes from DNA of this quality has not been previously reported and suggests that this method can be applied to forensic DNA samples. This would allow accurate inferences of genetic relationships between samples.

We further processed 100 bp fragmented samples at lower input amounts. We investigated alternatives to Illumina’s WGA step by creating sequencing libraries using a standard library preparation technique and a high-sensitivity single-stranded approach (Gansauge, M.T. and Meyer M. 2013. Nat. Protoc. 8(4), 737-48). We report on sequencing libraries as an alternative input for Illumina Beadchips.

This research was supported by DOJ award 2012-DN-BX-K037 and University of Utah Research Foundation and Interdepartmental grants to LBJ.
1680W Computational pipeline for whole genome sequencing data analysis -- an application to trio families with 22q11 deletion. J. Call, K. Coleman, Z. Zhang, B. Morrow. Genetics Dept, Albert Einstein College of Medicine, New York, NY; 2) Human Genetics Dept, Children’s Healthcare of Atlanta and Emory University School of Nursing, Atlanta, GA.

Whole genome sequencing (WGS) has become a viable strategy to discover variants associated with human disease. To facilitate WGS data interpretation, we established an automated pipeline for genetic variant (SNVs, INDELS) identification and annotation, using BWA, PICARD, GATK and ANNOVAR. For nonsynonymous SNPs, we predict the effect of amino acid changes, by combining the output of BLOSUM62, SIFT and Polyphen2. We derive the MAF of known variants from the 1,000 Genomes Project or the NHLBI Exome Sequencing Project. We also utilize multiple tools including CNVnator, ERDS, and BreakDancer, which consider both depth and split-read information to generate CNV calls. To obtain a reliable set of CNVs, we follow up CNVs that are reported by at least two tools. We annotate known CNVs according to the Database of Genomic Variants. We applied this approach to WGS data from two unrelated families in which the child has velo-cardio-facial/DiGeorge/22q11.2 deletion syndrome and both parents are normal. One child has a heart defect (tetralogy of Fallot; TOF) and the other has a normal heart and aortic arch. The known 22q11.2 deletions were verified by WGS. We identified 3,372,285 SNVs and 762,680 INDELS per sample. A total of 98% and 80.9% of them, respectively, are previously known SNPs found in dbSNP. We phased SNVs and INDELS using both trio and alignment information. This approach allows us to aggregate single base-pair level variants into gene level and dissect compound mutations within one gene. After phasing, we focused on highly possible deleterious variants, including nonsynonymous SNVs that at least one program predicted to be deleterious, stop gain/loss, splicing SNVs and frame-shift INDELS. We also filtered them with MAF <1% from the 1,000 Genomes Project. For each child, we counted the number of deleterious haplotypes for each gene and identified 20 candidate genes with two deleterious haplotypes in the child with TOF but not in the control. As we previously collected 303 heart related genes from literature, we examined whether these genes in the child with TOF were enriched with deleterious haplotypes compared to control, but resulted in no significant differences. We then checked the other 22q11.2. We did not find any highly possible deleterious variants. For CNVs, we detected 1,256 novel deletions and 215 novel duplications per sample. Overall, this approach can be applied to large datasets to identify genetic modifiers of 22q11DS.

1681T Blueprint: Resources provided by the large-scale Epigenomics project. L. Clarke, D. Richardson, S. Wilder, P. Flicek, The Blueprint Consortium. Vertebrate Genomics, European Molecular Biology Laboratory, European Bioinformatics Institute, The Wellcome Trust Genome Campus, Cambridge, United Kingdom.

Blueprint is a large-scale epigenomics project funded through the EC Framework Program 7 as the European Union’s entry in the International Human Epigenome Consortium (IHEC). Blueprint aims to provide epigenomic data on several blood cell types both normal and diseased to help the scientific community improve its understanding of haemopoiesis. This poster described the pipeline used to analyse the sequence data generated by our experiments and how the community can access the data. Blueprint is generating RNA-seq, ChIP-seq, DNase1-seq and WGBS-seq data on several different blood cell types. This sequence data is aligned to the genome and then appropriate signal calling analyses for the data type is undertaken. For the RNA-seq data expression levels are quantified, peaks are called for the ChIP-seq and DNase1-seq data and the WGBS-seq has hypo and hyper-methylated regions identified. We also perform some integrated analyses that include differential expression analysis and functional analysis of the histone marks using ChromHMM. Blueprint releases new data regularly. On data release the raw sequence data and analysis alignments associated with the release are made available via the European Genome-phenome Archive (EGA), to users given permission by the Blueprint Data Access Committee (DAC). There are three primary mechanisms for release of the non-unique data types, such as expression levels, chromatin states and methylation levels. Our ftp site: (ftp://ftp.ebi.ac.uk/pub/databases/blueprint) holds the raw files and a Track Hub that can be associated with both Ensembl and UCSC Genome Browser to allow easy visualisation of the data. This poster will primarily focus on these resources. We also provide a BioMart instance (http://blueprint.bsc.es/) to allow users to query the data and a Genomatix Browser (http://blueprint.genomatix.de/) for data visualisation. This poster gives an overview of the Blueprint project as well as providing details of our analysis pipelines and information about data availability and accessibility for the project.

1682F Exome sequencing identifies de novo mutations in patients with intellectual disability and epilepsy. J. Halvardson, A. Zaghlool, AC. Thuresson, L. Feuk, Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

Exome sequencing has proven to be an effective method to identify disease genes and have lead to the identification of causative genes in a large number of syndromes in recent years. Here we present the results of exome sequencing on 25 trios on a subset of patients with intellectual disability (ID), with additional symptoms of either epilepsy (19 trios) or congenital heart defects (6 trios). All trios have a negative family history and the parents are non-consanguineous. All patients have previously been screened for copy number aberrations using arrays, without clinically relevant findings.

We can report validated de novo single nucleotide variants (SNVs) or indels in coding regions in 11 of the 25 trios (44%). Of the 11 genes with validated de novo mutations, only two have previously been associated with ID or epilepsy (SCN2A and KCNA1). Thus, a high number of our findings are potentially novel disease genes. A subset of the genes have mouse knockout data that mirrors the patient phenotype, supporting a causative role for the de novo mutations. In one patient with severe ID we found a mutation in a highly conserved chromatin remodelling factor, which we are now following up with functional studies. This adds to a growing list of genes involved in chromatin remodeling that have been linked to intellectual disability and neurodevelopmental syndromes. The majority of the validated SNVs were missense mutations of conserved nucleotides in genes expressed in the brain. Several genes are involved in basic cell functions and we are currently investigating the possibilities of performing RNA sequencing on peripheral blood cells from trios to find patterns of expression in patients compared to healthy controls.

During this work we have created an in-house database of all sequenced exomes. This database contains several SNVs seemingly unique as compared to dbSNP, but common in the Swedish population. For example, one SNV primes to a coding difference in SCN2A and is common in our study. Thus, a high number of our findings are not occurring in the dbSNP database. Using this database as a filtering step leads to a significant reduction in the number of false positive SNV calls. Our data highlights the importance of a population specific filtering database when doing trio sequencing.

1683W An ensemble genotyping approach for whole genome sequencing to reduce erroneous variant calls. I.H. Lee, J.H. Park, M.B. Neu1, K. Lee2, T. Hambuch, I.S. Kohane1,2, R.C. Green1, S.W. Kong1, The MedSeq Project. 1) Boston Children’s Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Illumina, Inc., San Diego, CA; 4) Brigham and Women’s Hospital, Boston, MA.

Whole-genome sequencing (WGS) has been effective in identifying causal genomic variants of rare and common complex diseases as well as developing personalized treatment strategies. Accurate genotyping is essential to utilize genome sequence information in clinical settings as aimed by the MedSeq Project, a randomized clinical trial for integrating WGS in medicine. The current consensus is to use sequencing using multiple platforms to achieve the highest level of accuracy, then to validate all disease-associated variants for a clinical report using Sanger sequencing. In the course of exploring the plausibility of reducing erroneous variant calls in clinical settings, we deployed an ensemble of 8 variant calling algorithms to 17 publicly available paired WGS samples prepared using 2 different platforms. Our ensemble approach significantly reduced possible false positives by 48.8% for single nucleotide variants (SNVs), 50.5% for loss-of-function SNVs, 77.2% for small indels, and 96.4% for de novo mutation (DMN) candidates when we assumed higher false positive rates among platform-specific variants. Among 10 independently validated DNMs in the same trio dataset, 8 somatic or germline mutations were accurately detected by our proposed method compared to none by a pedigree-aware variant calling algorithm. Moreover, we could increase the sensitivity for detecting known disease-causing mutations by combining the results from each algorithm. Our results suggest that the ensemble of multiple variant calling algorithms is a cost-effective way to reduce both false positive and false negative variant calls. The source code of our ensemble genotype calling approach is freely available to the academic research community.
1684T Investigating the significance of genetic proximity for HLA matched donors and recipients in unrelated allogeneic stem cell transplantation. A. Madbouly1, V. Paunic2, F. Maier1. Genomics Research, National Mar- donor Program, Minneapolis, MN.

Survival after hematopoietic cell transplantation is dependent on HLA matching between donors and recipients, ethnicity and other factors, most of which are unknown. Donors and recipient ethnic origin is usually self-reported, and it is not clear if matching based on genetic ancestry would result in better survival. While an individual’s genetic composition is relatively fixed, self-identified race and ethnicity (SIRE) is a result of self-perception and can change over time. We aimed to investigate the effect of multiple measures of pairwise donor/recipient (D/R) genetic distance, using ancestry informative markers (AIMs), on transplant outcome. The AIMs panel selected for this pilot study consists of 500 autosomal SNPs. DNA samples from 300 fully HLA matched D/R pairs were genotyped for AIMs and were initially clustered using Structure and the 1000 Genomes dataset for reference parental populations. A genetic vector distance was estimated based on the distance between the D/R Structure vectors (describing individual ancestral proportions) in Euclidean space. Additionally, principal component analysis (PCA) was performed on the D/R genotypes for mapping genetic variation and Euclidean distance between D/R pairs was calculated on the PCA map to reflect genetic proximity. While a plateau was not explicitly reached, the optimal number of Structure clusters was estimated at K=5 based on the ad hoc delta(K) measure. The average Euclidean distance between donor and corresponding recipient Structure vectors was 0.347 (range [0.001, 1.376], p = 1). For the PCA, the first, second and third PCs accounted for 3.7%, 1.5% and 0.9% of the total variance respectively. The average PCA distance based on the first three PCs was estimated to be 1.4 (range [0.017, 9.18], p = 0.02), 2.05 (range [0.06, 10.3], p = 0.29) and 2.66 (range [0.19, 10.52], p = 0.786). Additionally, the average PCA distance was higher for SIRE mismatched D/R pairs (for the first PC, mean = 2.6, range [0.06, 9.2], p = 0.015) than SIRE matched D/R (mean = 1.14, range [0.01, 5.9], p = 0.99). Further analysis is being conducted on Caucasian sub-populations within the studied D/R cohort and a multivariate transplant outcomes analysis is underway. A power analysis will be performed to determine whether to proceed with the next phase of this study.

Genetic ancestry based analyses are expected to decrease the effect of confounding factors associated with self-reported ethnicity.

1685F Heterogeneity of global gene expression microarray designs in detecting differentially expressed genes. G. Malera1, D. Noei2, A. Ferrarini2, L. Xumerle1, V. Mijatovic1, P.F. Pignatti1, M. Delledonne2. 1) Department of Life and Reproduction Sciences, University of Verona, Verona, Italy; 2) Biotechnology Department, University of Verona, Verona, Italy.

Microarray technology is widely used for gene expression studies by many laboratories worldwide. Microarrays vary for the type of substrate used and for the type and number of oligonucleotide probes implemented. Moreover a disparate list of procedures to subtract background noise (BS) and normalize data (DN) among samples is available. All these factors help to make these reliable tools quite heterogeneous and heterogeneity may play an important role in identifying or not the differentially expressed (DE) genes when conducting global gene expression studies. To address this issue we essayed 4 different microarray platforms to analyze 2 Vitis vinifera berry developmental stages. Microarray data were processed using 20 different BS-DN combinations. The same RNA samples were also analyzed by RNASeq technology, whose results were used as reference values. Microarray performances in detecting DE were assessed using several measures including correlation between fold-change, classification functions and the area under curve (AUC) of receiver operating characteristic (ROC) curves. The number of DE genes changed from one microarray design to another. Although arrays are commonly deemed reliable tools for gene expression studies their performances are not all the same when conducting studies of differential gene expression.


Genetics professionals need access to organized information about reported relationships among sequence variation, genes, and phenotype to translate advances in genomic technology into medical practice. This presentation summarizes how ClinVar, Gene, MedGen and related tools interoperate to centralize the data, standardize its representation, and facilitate access.

ClinVar aggregates and archives information about sequence variation and assertions made about its clinical significance. ClinVar reports the data in multiple ways to support both interactive queries and integration into workflows. For example, clinicians and researchers may search a DNA or protein location to learn what is known about the clinical significance of variation, both now and previously. Users can access the evidence for or against an asserted phenotype based on its level of expert curation, and how an interpretation may have changed over time. Expert panels and professional societies can also submit assertions about pathogenicity, e.g. in a practice guideline. Aggregation and centralization of variant data from diverse sources such as testing laboratories, genome-wide research and curatorial groups will reduce the burden on professionals who no longer need collate this information independently and enable them to perform an informed and comprehensive assessment.

NCBI’s Gene database organizes information that defines or is related to a gene: sequence, phenotypes, citations, pathways, interactions, and regulatory aspects. The gene sequence section contains both sequenced portions and RefSeqGene, reports the sequence standards against which variation is reported. The phenotype section summarizes the names of disorders related to a gene, and information about dosage sensitivity. The pathways and interactions sections report other genes with products that, based on their interactions, may contribute to the same phenotype.

MedGen organizes information about conditions such as clinical features, related genes, practice guidelines, ontologies, GeneReviews, OMIM, published literature, concise resources, and tests registered in the NIH Genetic Testing Registry (GTR). Synonyms and term hierarchies support querying and navigation to related information. The advanced search option allows retrieval based on combinations of clinical features and other concepts, including existence of related sequence variation.

1687T Meta-analysis of FMRP mRNA Target Datasets Reveals Highly Associated mRNAs Mediated by G-quadruplex Structures Formed by Clustered WGGGA Sequences. J. Suhl1, P. Chapra1, S. Warren1,2,1) Dept. of Human Genetics, Emory University, 615 Michael St, Atlanta, GA; 2) Dept. of Biochemistry and Pediatrics, Emory University, Atlanta, GA.

Fragile X syndrome (FXS), a common cause of inherited intellectual disability and a well known monogenic cause of autism spectrum disorder (ASD), is the result of loss or dysfunction of the selective RNA-binding protein FMRP. Since FMRP appears to modulate activity-dependent translation of its target mRNAs at the synapse, a major research effort has been to identify these mRNA targets. Three large-scale studies have attempted to characterize the mRNAs bound by FMRP, each using different methods and each generating lists of putative target genes, leading to distinct hypotheses by which FMRP recognizes its targets either by RNA structure or RNA sequence. However, very little in depth analyses have been performed to identify the level of consensus among the studies. Here, we describe a meta-analysis of these three large-scale FMRP target studies. Consensus lists comprised of mRNAs found to consistently interact with FMRP across the independent methods were generated, conferring a higher likelihood of bona fide association. Additionally, we examined all datasets for sequence elements within the target RNAs to validate the recently reported FMRP binding sequences ACUK and WGGGA. We discovered that the sequence WGGGA was significantly enriched in multiple FMRP target datasets, supporting this motif as a strong candidate recognition sequence. Unlike WGGGA, the ACUK pattern was not enriched in the datasets suggesting that this motif may not be directly recognized by FMRP. The strong enrichment of WGGGA in FMRP targets, coupled with the previous suggestion that G-rich secondary structure serves as a recognition element, indicates that the motif may be able to form a G-quadruplex secondary structure. We observed that the WGGGA motifs in the FMRP targets are highly clustered and are consistent with the requirements for the G-quadruplex configuration, thus implicating the involvement of G-rich secondary structure as recog- nition elements for FMRP. Our analysis identifies the mRNA targets most likely affected by the loss of FMRP and, consequently, the genes critical for neuronal development and synaptic plasticity.
1689F
Real-world performance of five long-range PCR enzymes to amplify ~10kb amplicons from human genomic DNA. H. Jia, K. Wang. Zilka Neurogenetic Institute and Department of Psychiatry, Keck School of Medicine, University of Southern California, Los Angeles.

Long-range PCR remains a flexible, fast, efficient and cost-effective choice for sequencing candidate genomic regions or genes, especially when combined with personal genome sequencers such as GS-20 and Ion Torrent. Several long-range DNA polymerases are available commercially, and all of them are advertised as being able to amplify up to 15kb or longer genomic DNA. Although it is likely that they may all work well for specific genomic regions under highly optimized conditions, it is unclear what their real-world performance on randomly chosen amplicons in human genomic DNA can be. Knowing the different characteristics of these enzymes help select enzymes for use in next-generation sequencing experiments. In the current study, we evaluated five long-range DNA polymerases to amplify three amplicons, with sizes of 5.8kb, 9.7kb, 12.9kb and Tm values of 54.5 °C, 63.5 °C and 54 °C respectively. These five long-range polymerases include Invitrogen SequaQ Long PCR and AccuPrime Taq DNA Polymerase, TaKaRa PrimeSTAR GXL DNA Polymerase, KAPA long Range HotStart DNA polymerase and QIAGEN LongRange PCR Polymerase. The advertised lengths of PCR products for these enzymes are 15kb, 20kb, 30kb, 20kb and 40kb, respectively. All experiments were designed according to the reaction mixture and cycling conditions of the manuals. We found that PrimeSTAR can amplify all the targets with good quality using two-step unified PCR conditions. SequaQ Long and AccuPrime can amplify 5.8kb and 12.9kb but not 9.7kb amplicon, unless PCR conditions are optimized. KAPA can amplify 5.8kb but not the two larger amplicons using their recommended long targets (5kb-19kb) PCR reaction and cycling conditions. QIAGEN cannot amplify any of the three amplicons according to the manual’s conditions for very long-range PCR. In practice, many amplicons may need to be amplified at the same time, but primer pairs for them may have very different Tm values. Therefore, we also varied the annealing conditions for selected enzymes, and evaluated how sensitive the enzymes are to alterations of the annealing temperature and extension time. The results indicate that PrimeSTAR polymerase can amplify amplicons with different sizes and Tm values using generalized enzyme conditions for most PCR conditions to obtain optimal performance. Our results provided a useful guide to other researchers working on next-generation sequencing of candidate genes or regions by long-range PCR reactions.

1690T

With an increasing interest in exome analysis for research and clinical diagnostics the number of exome capture kits are available on the market is increasing. The two most published and commonly used are the Agilent SureSelect and NimbleGen SeqCap EZ kits. More recently Life Technologies released their TargetSeq kit for use on the Ion Proton instrument. NimbleGen has adapted their kit for use on the Ion Proton, while Agilent has yet to release an Ion-compatible SureSelect kit. Therefore, we evaluated the Life Technologies TargetSeq and NimbleGen SeqCap EZ v3 kits for exome sequencing on the Ion Proton. Both kits are liquid-phase capture kits, contain about 2 million capture probes and cover greater than 20,000 of the known coding genes plus varying amounts of RNA genes, miRNA binding sites and UTRs. Due to differences in probe design, TargetSeq has a target region of 46.2Mb whereas the SeqCap kit targets 64Mbases. We compared the performance of these kits on the Ion Proton in terms of capture ability and specificity, coverage of the intended targets and coverage density sufficient for mutation analysis. Exome captures were performed according to the manufacturer’s instructions followed by sequencing on the Ion Proton with the P1 chip. Both kits were able to capture their intended targets fairly similarly with about 85% of sequence reads being on target and both kits covered more than 92% of their targets with 1x coverage or greater. The remaining target sequence is either not well captured or there is no PCR bias against them. In order for exome sequencing to reliably identify mutations the amount of target covered with at least 20x or higher is important. In this respect, the two kits differ. With the TargetSeq we can sequence greater than 86x with 20x coverage. The SeqCap runs consistently with significantly lower coverage. The difference in quality of coverage is likely due to the fact that the SeqCap kit covers almost 40x more bases than the TargetSeq kit, therefore lowering the average coverage per base. Given the need for sufficient coverage, and the fact that current Ion Proton output is limited by the P1 chip, the TargetSeq kit is preferable. However, since the other performance characteristics are similar, the SeqCap kit can be a valuable kit when sequence output on the Proton instrument is improved in the future.

1691F
A single-tube high-plex PCR approach for targeted massively parallel sequencing applied to FFPE-tumour derived material. T. Nguyen-Dumont1, B. J. Pope1,2, F. Hammett, M. C. Southey1, D. J. Park1. 1) Genetic Epidemiology Laboratory, Department of Pathology, Medical Building, The University of Melbourne, Victoria 3010, Australia; 2) Victorian Life Sciences Computation Initiative, The University of Melbourne, 187 Grattan Street, Carlton, Melbourne, Victoria 3010, Australia; 3) Department of Computing and Information Systems, The University of Melbourne, Victoria 3010, Australia.

High-plex PCR, as a sequence enrichment method for massively parallel sequencing (MPS), has been made available recently through a range of commercial technologies, with costs ranging from ~100 to ~1000 per sample. However, protocol complexity, expense, and limited design flexibility, reduce their suitability in settings involving modest target size, or requiring low cost and high-throughput. In addition, methods to sequence DNA extracted from formalin-fixed, paraffin-embedded (FFPE) material have achieved good success so far. To address these limitations, we have developed Hi-Plex, a PCR-MPS strategy intended for high-throughput screening of multiple genomic target regions. Our library-building approach integrates a simple, automated primer design process to enable high-throughput MPS. As a result, Hi-Plex achieves minimal size, feature, and sequence permutations permits, inclusive thermocycling conditions and ‘clamp’ amplification bias reduction, our protocol is simple due to its single-tube nature. It is also cost and time-effective, using readily available reagents, is independent from expensive instrumentation, and requires minimal optimisation.

In a 60-plex test setting, targeting the breast cancer predisposition genes PALB2 and XRCC2, Hi-Plex applied to 100 ng LCL-derived DNA resulted in 93.3% (56/60), 98.33% (59/60), and 100% of targeted amplicons representing 5-fold, 10-fold and 12.5-fold of the mean, with 97-98% of reads on-target. In order for exome sequencing to reliably identify mutations the amount of target covered with at least 20x or higher is important. In this respect, the two kits differ. With the TargetSeq we can sequence greater than 86x with 20x coverage. The SeqCap runs consistently with significantly lower coverage. The difference in quality of coverage is likely due to the fact that the SeqCap kit covers almost 40x more bases than the TargetSeq kit, therefore lowering the average coverage per base. Given the need for sufficient coverage, and the fact that current Ion Proton output is limited by the P1 chip, the TargetSeq kit is preferable. However, since the other performance characteristics are similar, the SeqCap kit can be a valuable kit when sequence output on the Proton instrument is improved in the future.
1692W
PhenoTips: Patient Phenotyping Software for Clinical and Research Use. M. Girdea1,2, S. Dumitriu3, M. Fiume1, S. Bowdin1,4,5, K. Boycott1, S. Chénier1, D. Chitayat1,4,8, H. Faghfoury9,10, M.S. Meyn1,3,4,5,7,11,12, P.N. Ray1,3,4,5,12,13, J. So1,4, D.J. Stavropoulos13,16, M. Brudno1,2,11. 1) Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; 2) Centre for Computational Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada; 6) Children’s Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada; 7) Centre Hospitalier Universitaire de Sherbrooke, Quebec, Canada; 8) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, Toronto, Ontario, Canada; 9) Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 10) University Health Network, University of Toronto, Toronto, Ontario, Canada; 11) Genetics and Genome Biology Program, Hospital for Sick Children, Toronto, Ontario, Canada; 12) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 13) Department of Paediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 14) Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 15) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, Ontario, Canada.

We have developed PhenoTips, a deep phenotyping tool and database, specifically designed for phenotyping patients with genetic disorders. Our tool closely mirrors clinician workflows so as to facilitate the recording of observations made during the patient encounter. Phenotypic information is represented using the Human Phenotype Ontology; however, the complexity of the ontology is hidden behind a user interface, which combines simple selection of common phenotypes with error-tolerant, predictive search of the entire ontology. The software provides a series of features that help reduce the clinician’s workload during the clinical examination. Together with standardized phenotypic data, PhenoTips supports entering demographic information, medical history (including prior laboratory results), family history, various measurements, relevant images depicting manifestations of the patient’s disease, genetic tests and their results, as well as additional notes for each of these categories. A pedigree drawing tool which enables the collection of advanced family histories is currently under development. The software automatically plots growth curves for a variety of measurements, selects phenotypes reflecting abnormal measurements, instantly finds disease GEnes) Canada project (http://care4rare.ca/), and the Undiagnosed Disease Program at the NIH. Our source code and a demo version of PhenoTips are available at http://phenotips.org.

1693T

Single-cell analysis by using PCR and array assays is well established for determining aneuploidy, CNV, and genotyping single cells. Next-gen sequencing of single cells presents many barriers to complete and reproducible analysis, including incomplete genome coverage and irreproducible results. In some applications the information from multiple single cell experiments can be combined to increase coverage and reproducibility. However a surprising number of applications can be successfully executed using partial coverage of the genome as long as the coverage is reproducible. These technical applications include a) aneuploidy and copy number variation determination, b) SNP or other single nucleotide variations in a fraction of the genome, and c) identification of complex populations of cells. These technical applications are important for commercial applications such as pre-implantation genetic screening and diagnosis (PGS and PGD), prenatal diagnostics from single or small numbers of fetal cells in maternal circulation, cancer diagnostics from circulating tumor cells, and identification of infectious disease. In these applications coverage can be compromised as long as the partial coverage is reproducible. We have sequenced human and bacterial cells using a version of the Rubicon PicoPLEX technology that is being developed as a single-cell NGS library kit. This PicoPLEX-scD single-cell NGS prep is as simple as the PicoPLEX WGA kit, which is currently used for microarray and PCR studies and diagnostics from single cells. The PicoPLEX-scD prototype kits were used for the above technical applications. Sequencing quality, genome coverage, and reproducibility were measured in flow sorted and microdissected human cells. To verify that the MiSeq NGS could be used for PGS/PGD applications, as many as 48 single human cells were multiplexed on a single lane. Megabase losses or gains of copy number were reproducibly measured with as few as 200,000 clusters per sample. Partial genotyping and variant identification using single cells were also measured. Finally, single cells were studied in mixtures of other genomes. The results show that reproducible incomplete coverage was achieved by NGS of PicoPLEX libraries, and that many of the requirements for diagnostics of single cells could be achieved with minimal numbers of sequencing reads.
Cis and trans protein quantitative trait loci (pQTLs) identified using a high-throughput protein assay in 297 individuals from the AddNeuroMed cohort. The European collaboration for the discovery of novel biomarkers for Alzheimer's disease. J.E. Mollon1,2, M. Sattler1,2, S. Kiddie1,2, C. Johnstone1,2, K. Lunn1,2, P. Proitsi1,2, J. Powell1, A. Hodges1,2, S.K. Nelson1, A. Stewart1, S. Williams3, H. Soininen1, I. Kloszewksi1, P. Meccocci1, M. Tsalaki5, B. Veillas6, S. Lovestone1,2, S. Newhouse1,2, R. Dobson1,2. 1) Institute of Psychiatry, King's College London, London, United Kingdom; 2) NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation Trust; 3) SomaLogic, Boulder, Colorado, United States of America; 4) Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 5) Medical University of Lodz, Lodz, Poland; 6) Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy; 7) 3rd Department of Neurology, Aristotle University, Thessaloniki, Greece; 8) INSERM U 558, University of Toulouse, Toulouse, France.

There has been much success in identifying gene expression quantitative trait loci using high-throughput gene expression measures such as array-based methods or sequencing (RNA-Seq), but high-throughput methods for quantifying proteins are not as widely available. In this study we use an aptamer-based protein assay from SomaLogic which quantifies over a thousand proteins from a single sample. Samples from 106 individuals with Alzheimer's Disease (AD), 90 with mild cognitive impairment (MCI) and 101 healthy elderly controls were selected from the AddNeuroMed study. Genotyping and imputation resulted in 6,345,198 SNPs for analysis, and a total of 1016 aptamers targeting 1001 proteins were assayed in plasma from the same individuals. We applied regression models using QUICKTEST, measuring single-SNP effects after adjusting for age, gender and 5 genetic principle component axes. A Bonferroni threshold was applied (p<4.93x10^-14), adjusting for multiple SNPs and proteins. Significant associations were further modelled in R, adjusting for original covariates as well as disease status and a single PC axis derived from the protein data. New associations with 87 proteins were discovered, and we replicated results from a previous study with overlapping samples. Twelve of the new associations were 'cis', defined as SNPs within the coding region of a gene associated with that gene's protein product. A further 14 proteins were associated with SNPs outside the protein-coding gene but on the same chromosome.

Sixty-one proteins had associations with SNPs on different chromosomes ('trans'). The top cis association was between a SNP and rs319795 (3.83x10^-7), a SNP in the third intron of MST1, the gene encoding MSP. This SNP is predicted damaging in an alternate transcript of MST1, and associated with Crohn's Disease, ulcerative colitis and primary sclerosing cholangitis. The top trans association was between MMP8 and rs12614 (p=2.8x10^-70), a non-synonymous (predicted damaging) SNP in FB112 (an alternate splice variant of CFB). We will describe properties of variants we have found to associate with protein plasma levels and highlight those previously identified as being disease related through GWAS studies. We conclude that the study of genetic effects on intermediary phenotypes such as proteins yields promising results, and may allow us to begin to unravel the mechanisms through which genetic effects manifest in disease.
Mixed model approaches for transcriptome profiling of reciprocal dosage imbalances. A. Ragavendran1, I. Blumenthal1, S. Erdin2, L. Kiel1, K. Roeder3, B. Devlin1,4, M.E. Talewski1,2,3, 1 Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2 Departments of Neurology and Genetics, Harvard Medical School, Boston, MA; 3 Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 4 Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 5 Department of Statistics, Carnegie Mellon University, Pittsburgh, PA.

Methods for analysis of RNA-Seq data are still in a nascent stage of development. Current packages for analyzing differential expression (DE) are developed primarily to accommodate low replication and simple designs. Few methods have been established to model family-based designs. We present some insights from ongoing statistical analysis of RNA-Seq count data associated with reciprocal dosage imbalances, which collectively represent some insights from ongoing statistical analysis of RNA-Seq count data. Inference for GLMMs is an ongoing area of research and we explore randomization procedures to generate empirical significance due to copy number state. Inference for GLMMs is an ongoing area of research and we explore randomization procedures to generate empirical significance due to copy number state. The second is to fit contrasts within a cell-means model to delineate effects due to age imbalance in the genome. The first is to evaluate the effects of deletion/duplication using a regression on the copy number for linear trends in DE.

The second is to fit contrasts within a cell-means model to delineate effects due to age imbalance in the genome. The first is to evaluate the effects of deletion/duplication using a regression on the copy number for linear trends in DE.

For statistical inferences in family-based designs, correlations due to shared genetic background are accommodated for family, while correlations between family members are modeled from pedigree-based estimation of relatedness. We show that two complementary GLMMs are a natural extension to GLMs and that it can potentially use other measures of regulatory activity (MNase coverage DHSs (mean=170, min=18, max=1195), that was consistent strong allele-specific bias in coverage (95.1% on average), within high-coverage DHSs (mean=170, min=18, max=1195). Nevertheless, we found that the FP estimates for a SNP were well correlated between the replicates where it was seen (mean r=0.88, sd=0.09). We focused on the set of SNPs overlapping more than 2 replicates (3.6% SNPs, n=1675, seen in > 2 replicates). This was expected that FP estimates for a SNP were well correlated between the replicates where it was seen (mean r=0.88, sd=0.09). We focused on the set of SNPs overlapping more than 2 replicates (3.6% SNPs, n=1675, seen in > 2 replicates). This was expected that it intersected with RSNP. There was large variation in the number of SNPs obtained from DNase-seq mapped DHS reads from five GM12878 lymphoblastoid cell lines replicates (R1-R5) for NA12878 released by the ENCODE Project. The read depth of the two alleles (from the DP4 annotation) at each SNP was used to estimated the FP for a total of 54533 variants (46652 distinct SNPs) that were called from DHSs-mapped reads from all replicates that intersected with RSNP. There was large variation in the number of SNPs called per replicate (mean=10807, sd=6065) and little overlap between replicates (3.6% SNPs, n=1675, seen in > 2 replicates). This was unexpected due to the short read lengths (20bp) and low read coverage at DHSs (mean=6.32, min=1, max=1195). Nevertheless, we found that the FP estimates for a SNP were well correlated between the replicates where it was seen (mean r=0.88, sd=0.09). We focused on the set of SNPs overlapping more than 2 replicates and observed significant (p<9.16x10^-7) FP estimates for 6.9% (n=116) after correcting for multiple testing. These SNPs showed a very strong allele-specific bias in coverage (95.1% on average), within high-coverage DHSs (mean=170, min=18, max=1195), that was consistent across five replicates. These results suggest that our FP approach is feasible and that it can potentially use other measures of regulatory activity (MNase and FAIRE-seq).
1700W
DeepSAGE Reveals Genetic Variants Associated with Alternative Polyadenylation and Expression of Coding and Non-Coding Transcripts.
D.V. Zheleva1, E. de Klerk1, H.-J. Westra2, A. Mastrokilas2, S. Amini3, Y. Atyurek4,5,6, R. Jansen7, B.W. Penninx8, J.J. Hottenga9, G. Willemsen9, E.J. de Geus9, D.I. Boomsma10, J.H. Veltink10, L.H. van den Berg10, C. Wil- mena11, J.T. den Dunnen12,5,6, G.-J.B. van Ommeren12, P.A.C. ‘t Hoen1, L. Franke1, 1) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands; 2) Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; 3) Leiden Genome Technology Center, Leiden, The Netherlands; 4) Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Depart- ment of Biological Psychology, Netherlands Twin Registry, VU University, Amsterdam, The Netherlands; 6) Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Utrecht, The Netherlands.

Many disease-associated variants affect gene expression levels (expression quantitative trait loci, eQTLs) and expression profiling using next generation sequencing (NGS) technology is a powerful way to detect these eQTLs. We analyzed 94 total blood samples from healthy volunteers with DeepSAGE to gain specific insight into how genetic variants affect the expression of genes and lengths of 3′-untranslated regions (3′-UTRs). We detected previously unknown cis-eQTL effects for GWAS hits in disease- and physiology-associated traits. Apart from cis-eQTLs that are typically easily identifiable using microarrays or RNA-sequencing, DeepSAGE also revealed many cis-eQTLs for antisense and other non-coding transcripts, often in genomic regions containing retrotroponson-induced elements. We also identified and characterized splice junctions that result in the alteration of alternative polyadenylation sites, thereby potentially influencing the stability of messenger RNAs (mRNA). We then combined the power of RNA-sequencing with DeepSAGE by performing a meta-analysis of three datasets, leading to the identification of many more cis-eQTLs. Our results indicate that DeepSAGE data is useful for eQTL mapping of known and unknown transcripts, and for identifying SNPs that affect alternative polyadenylation. Because of the inherent differences between DeepSAGE and RNA-sequencing, our complementary, integrative approach leads to greater insight into the molecular consequences of many disease-associated variants.

1701T

Center for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 2) Department of Urology, (SIUT), Karachi, Pakistan; 3) Department of Histopathology, (SIUT), Karachi, Paki- stan; 4) Department of Human Genetics, University of Health Sciences (UHS), Lahore, Pakistan.

Abstract: Interferon-gamma inducible protein-10 (IP-10) is a chemokine secreted by immune and non-immune cells. IP-10 and its receptor CXCR3 have been shown to participate in the immune response against kidney allograft. We aimed to investigate the association of urinary IP-10 levels with rejection episodes in renal transplant patients during the first year post transplant. A total of 206 samples were included in this study. Of these, 118 samples were taken from 93 renal transplant recipients with various histopathological findings of rejection (急性 cellular ACR, borderline BLR and acute vascular rejection AVR) and of no-rejection episodes (NAD). Additionally, rejection profiles were also available from 20 deceased donor kidney transplantations. Urinary IP-10 levels were quantified by using the Human CXCL10/ IP-10 Quantikine ELISA kit. The data were analyzed using the Statistical Package for Social Sciences (SPSS) and MedCalc software. Statistically significant differences in the urinary IP-10 levels were found between rejection and control groups (p<0.001) and rejection vs. NAD group (p<0.004) groups. Among rejection groups, the mean IP-10 levels for ACR, BLR and AVR were 256.8±52pg/mL, 196±110pg/mL and 113±54pg/mL respectively. Among rejection groups, the mean IP-10 levels for ACR, BLR and AVR were 256.8±52pg/mL, 196±110pg/mL and 113±54pg/mL respectively.

1702F
A General Statistical Framework for Identifying Genetic Variants of Clinical Significance. L. Ma, M. Xiong. University of Texas School of Public Health, Houston, TX.

Fast and cheaper next generation sequencing (NGS) technologies will generate unprecedentedly massive (thousands or even tens of thousands of individuals) and highly-dimensional (up to hundreds of millions) genomic and epigenomic variation data. In the near future, a routine part of medical record will include the sequenced genomes. A fundamental question is how to efficiently extract genomic and epigenomic variants of clinical utility which will provide information for optimal wellness and interference strategies. Traditional paradigm for identifying variants of clinical validity is to test associ- ation of the variants. However, significantly associated genetic variants may or may not be usefulness for diagnosis and prognosis of diseases. Alternative association studies for finding genetic variants of predictive utility is to systematically search variants that contain sufficient information for pheno- type prediction. To achieve this, we introduce concepts of sufficient dimen- sion reduction (SDR) and coordinate hypothesis which project the original high dimensional data to very low dimensional space while preserving all information on response phenotypes. We then formulate clinically significant genetic variant discovery problem into sparse SDR problem and develop algorithms that can select significant genetic variants from up to or even tens millions of predictors with the aid of dividing SDR for whole genome into a number of sub-SDR problems defined for genomic regions. The sparse SDR is in turn formulated as sparse optimal scoring problem, but with penalty which can remove row vectors from the basis matrix. To speed up computation, we apply the alternating direction method for multipliers to solving the sparse optimal scoring problem which can easily be implemented in parallel. To illustrate its application, the proposed method is applied to simulation data and the NHLBI’s Exome Sequencing Project (ESP) dataset as well as the TCGA dataset.

1703W
The n=1 problem in human genetics: identification of rare disease mutations from single genomes. A.B. Wilfert1, J.N. Constantino2, D. Conrad1, 1) Genetics, Washington University in St. Louis, St. Louis, MO; 2) Psychiatry, Washington University in St. Louis, St. Louis, MO.

Autism and azoospermia are examples of reproductively lethal diseases with a large population prevalence (about 1%) and extensive genetic hetero- geneity (at least 500-1000 genes each). Standard epidemiological methods have begun to fail in the genetic analysis of these traits as cohorts of realistic size will contain many individuals with private, causal mutations that cannot be pinpointed by GWAS or linkage. Likewise, it has been esti- mated that over 40% of Mendelian disorders remain undiagnosed, which is in part due to sample size limitations. The purest presentation of these challenges is called the n=1 problem: given a single genome with a sus- pected monogenic disease phenotype, can one identify the causal mutation? We have created the first statistical framework that explicitly addresses the n=1 problem. This is an integrative framework that relies heavily on published data sets and integrates gene physiology, functional protein-level models and patterns of population genetic variation to model disease and healthy varia- tion. We use variants from the Human Gene Mutation Database (HGMD) and the 1000 Genomes Project to train and test these models. Spike-in analysis shows that over 80% of the HGMD variants tested have been prioritized into the top 10 variants within a typical healthy exome and ROC curves show that this integrated prioritization method outperforms each of its component predictors, with an AUC of 0.990. For over 100 Mendelian diseases, each with an average of 5 distinct mutations in HGMD, our method correctly identifies the causal mutation as the most pathogenic event within an otherwise healthy exome, indicating that, in principal, our method solves the n=1 problem for these diseases. We describe applications of this method in a few disease cases and discuss the reproducible methods for disease and azoospermia, identifying candidate muta- tions in each of these cases, where the genetic etiology was previously unknown. We apply our method to describe the distribution of the most pathogenic mutation detected in 1000 healthy individuals sequenced by the 1000 Genomes Project. We discuss the complications of applying n=1 methodology in the presence of somatic mutations, which may be misclassified as germline variants when sequencing primary cells. While the number of developmentally acquired somatic mutations detectable in a given tissues may be small (e.g. 30-300) we show that these are much more likely to be appear pathogenic than germline variants by pathogenicity assessment tools.
1704T
Meta-analysis of twelve genome-wide association studies (GWAS) identifies novel genetic loci associated with mammographic density phenotypes. S. Lindström1, D. Thomas1, A.D. Pateron1 2,12, D. White2,4, E. Gierach2, J. Stone6, A.A. Douglas1, I. dos Santos Silva5, J. Benitez3,4, C. Scott1, P.A. Fasching11, L. Baglietto11, M. Southey4, G. Giles12, M. Pollan4, J. Figueroa5, F.J. Couch11, J.L. Hopper1, P. Hall5, D.F. Easton11, N.F. Boyd12, M. Vachon12, R.M. Tamimi5 8, 10, and S.E. Hopper12 11. Markers of Density (MODE) consortium. 1) Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School Of Public Health, Boston, MA, USA; 2) Centre for Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 3) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 5) Hormonal and Reproductive Epidemiology Branch, National Cancer Institute, Rockville, MD; 6) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, The University of Melbourne, Melbourne, Australia; 7) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI; 8) Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK; 9) Centro Nacional de Investigaciones Oncologicas, Madrid, Spain; 10) Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; 11) Erlangen University Perinatal Center, Department of Gynaecology and Obstetrics, Erlangen University Hospital, Friedrich Alexander University Alexander von Erlangen-Nuremberg, Erlangen, Germany; 12) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia; 13) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; 14) Campbell Family Institute for Breast Cancer Research, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; 2) Oncology and Division of Network Medicine, Brigham and Women’s Hospital, Boston, MA.

Mammographic density reflects the proportion of stromal and epithelial tissues in relation to adipose tissue in the breast and is one of the strongest risk factors for breast cancer susceptibility. Two studies selected women based on the extreme categories of mammographic density (one using percent density and the other dense area) and defined by the area of a mammogram that consists of stromal and epithelial tissue (dense area), adipose tissue (non-dense area) and the proportion of stromal and epithelial tissue in relation to total breast area (percent density). Therefore, on or more variations in genes could play critical roles in the diverse pathways which can further progress to breast cancer susceptibility loci. In summary, we report novel candidates at 4q13.3, 6q25.1, 10q21.2, 12q23.2 and 22q13.1, and a genome-wide significant association at 10q21.2. From our discovery phase, we identified novel genome-wide significant SNPs that were selected for replication in 5,000-8,300 women independent of families with participants recruited between 2006 and 2011. Digital ECG data were coded at the University of Glasgow ECG Core Laboratory. Exclusion criteria were use of QT prolongation medications at least two months prior to recruitment (n=619: 3.4%). QRSd<120 ms, atrial fibrillation or pacemaker. 13,722 individuals (5,082 families) were included in the analyses. Univariate and bivariate heritabilities were calculated using ASREML adjusted for age and sex. Results: The study population included 13,722 individuals (5,082 families) and the mean age was 47.7 years (range 20-81). We identified novel genome-wide significant (P<5x10^-8) loci for percent density at 4q23.2, 8p11.23 and 22q13.2, loci for dense area at 4q13.3, 6q25.1, 10q21.2, 12q23.2 and 22q13.1, and a locus for non-dense area at 8p11.12. Three of these regions have been identified as breast cancer susceptibility loci. In summary, we report novel associations between common genetic variation and three mammographic density phenotypes. Many of the genetic variants found to be associated with mammographic density are also associated with breast cancer supporting the hypothesis that there is a shared genetic basis between mammographic density and breast cancer risk.

1705F

Rationale: Cholesterol gallstone disease (CGD), one of the commonest digestive ailments in westernized and developing countries, is caused by a mixture of cholesterol flush and a failure to effectively remove it. Steinberg showed that cholesterol gallstone disease involves complex interactions among environmental and genetic factors but apart from female gender and increasing age, very little is known about the genetic interactions underlying the phenotypic expression of cholesterol gallstone disease. Therefore, on or more variations in genes could play critical roles in the diverse pathways which can further progress to cholesterol crystallization in gallbladder. In the present study, we performed multigene interactive analysis as genotyping score, Multifactor dimensionality reduction (MDR) and Classification and Regression Tree analysis (CART) to identify combinations of alleles among the hormonal, hepatocanalicular transporter and adiopogenesis differentiation pathway genes in modifying the risk for CGD. Design: The present study recruited a total of 450 subjects including 230 CGD patients and 220 healthy controls who completed an interview and provided blood. The study was approved by the local ethical committee and we followed the norms of Declaration of Helsinki. We analyzed common ESR1, ESR2, PGR, ADRB3, ADRB2, ABCG8, SLCO1B1, PPARgamma, and SREBP2 gene polymorphisms to find out combinations of genetic variants contributing to CGD risk. Genotyping was carried out by PCR-RFLP, ARMS-PCR and Taqman Assay. Statistical analysis was performed by using SPSS ver16 Results: Single locus analysis by logistic regression showed positive association of ESR1 IVS1-C397T (rs2234693), IVS1- A351G (rs9340799), PGR Ex17 C150T (rs11045819), PGR Ex17 -110CC (rs11045818), ABCG8 G145C (rs11875341), SLCO1B1 C Exon4A (rs11045819) and SREBP2 G1784C (rs2228314) with CGD risk. However, the MDR and CART analysis revealed ESR1 IVS1- C397T (rs2234693), ADRB3- T190C (rs499499) and ABCG8 G145C (rs11875341) polymorphisms as the best polymorphic signature for discriminating between cases and controls. The overall odds ratio for the applied multi-analytical approaches ranged from 4.33 to 10.05 showing an incremental risk for cholesterol crystal formation. Thus, the CART and MDR analysis showed the importance of ESR1 IVS1- C397T and ABCG8 G145C polymorphism in susceptibility to CGD risk Conclusion: Our multi-gene interactive analysis suggests that ESR1, ADRB3 and ABCG8 genetic variants confer significant risk for cholesterol gallstone disease.

1706W
Bivariate heritability analyses of cardiac conduction and repolarisation measures show a paucity of shared genetic influences. J. Alghamdi1, C. Hastie1, C. Schuz2, C. Brown1, L. Hocking2, M. Luciano3, D. Porteous4, A. Morris4, B. Smith4, A. Dominiczak2, L. Hocking2, S. Padmanabhan1. 1) College of Medical, Veterinary and Life Sciences, University of Glasgow, Institute of Cardiovascular and Medical Sciences, Glasgow, UK; 2) Aberdeen Pain Research Collaboration (Musculoskeletal Research), University of Aberdeen, Aberdeen, UK; 3) Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK; 4) Medical Genetics Section, Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; 5) Medical Research Institute, University of Dundee, Dundee, UK.

Genome wide association studies have shown predominantly concordant effects for SNPs associated with ECG conduction traits with shared physiologic processes (PR interval[PR] and QRS duration [QRSd]) and mostly discordant effects between traits that reflect different processes (QRSd and QT interval[QTc]). However, these SNPs explain a small fraction of the population variance. To estimate the proportion of shared genetic influences among ECG cardiac conduction and repolarisation measures we performed bivariate heritability in a large cohort of Scottish families. Methods: The Generation Scotland Scottish Family Health study is a nationwide cohort of families with participants recruited between 2006 and 2011. Digital ECG data were coded at the University of Glasgow ECG Core Laboratory. Exclusion criteria were use of QT prolonging medications at least two months prior to recruitment (n=619: 3.4%). QRSd<120 ms, atrial fibrillation or pacemaker. 13,722 individuals (5,082 families) were included in the analyses. Univariate and bivariate heritabilities were calculated using ASREML adjusted for age and sex. Results: The study population included 13,722 individuals (5,082 families) and the mean age was 47.7 years (range 20-81). We identified novel genome-wide significant (P<5x10^-8) loci for percent density at 4q23.2, 8p11.23 and 22q13.2, loci for dense area at 4q13.3, 6q25.1, 10q21.2, 12q23.2 and 22q13.1, and a locus for non-dense area at 8p11.12. Three of these regions have been identified as breast cancer susceptibility loci. In summary, we report novel associations between common genetic variation and three mammographic density phenotypes. Many of the genetic variants found to be associated with mammographic density are also associated with breast cancer supporting the hypothesis that there is a shared genetic basis between mammographic density and breast cancer risk.
A signal of polygenic inheritance from low frequency variants in case-control genome wide association studies elucidates genetic architecture of common diseases. Y. Chan1,2,3, E.T. Lim1,2,4, N. Sandholm1,2,5, A.J. McKnight6, 7, J.N. Hirschhorn7, S. Ripke2,4, M.J. Daly1,2,4, B.M. Neale2,4, R.M. Salem1,2,7, J.N. Hirschhorn1,2,3. 1) Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America; 3) Department of Endocrinology, Children’s Hospital, Boston, Massachusetts, United States of America; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Massachusetts, United States of America; 5) Folkhalsan Institute of Genetics, Folkhalsan Research Center, Biomedical Center, Helsinki, Helsinki, Finland; 6) Division of Nephrology, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; 7) Department of Biomedical Engineering and Computational Science, Aalto University, Espoo, Finland; 8) Nephrology Research Centre, Queen's University of Belfast, Belfast, United Kingdom.

In most complex traits and diseases, much of the heritability remains unaccounted for by common variants discovered from genome-wide association studies (GWAS). Thus it has been postulated that rare and low-frequency variants are likely to play a role in accounting for the remaining heritability. To test this hypothesis, we developed a novel method to test for polygenic inheritance specifically from low-frequency and rare variants from GWAS summary statistics. Our method utilizes the fact that there is more power to detect risk than protective variants especially for variants at lower frequencies. This would result in a higher than expected ratio of risk to protective (R/P) variants at a given P-value threshold. We tested our method on published GWAS results for various complex diseases and jointly analyzed, suggesting that although both phenotypes have distinct genetic components, they also share low frequency variants that contribute to a common genetic component in both diseases. Interestingly, while we observed a polygenic signal from low-frequency variants in both macroalbuminuria and ESRD, the signal was lost when the phenotypes were jointly analyzed, suggesting that the genetic variants underlying these two apparently related phenotypes may not overlap substantially, and that both sub-phenotypes should be studied separately. Thus, our method may help guide the design of future genetic studies of dichotomous traits and diseases.
**1709W**


Metabolic syndrome (MetS) is a combination of risk factors for increased adiposity/ obesity, glucose intolerance/ insulin resistance, dyslipidemia and high blood pressure. To account for non-alcoholic fatty liver disease as a hepatic expression of MetS, we performed a multivariate factor analysis to explore DNA variants associations with 20 interconnected traits in the Family Heart Study, including measures of liver attenuation (LA), alanine transaminase, risk factors for MetS, several measures of fat by CT imaging and a selected number of inflammatory markers. After selecting 133 top SNPs from mixed effects additive association tests with factor scores, the list narrowed further to 45 SNPs when they were tested with each single trait and using a threshold of p-values < 10^{-6}. Of them rs738409 of PNPLA3 associated significantly with LA (p=4.28E-25). Other candidates were also found. But there is no information available in predictive modeling with MetS candidates. We extend this work to predict based on genetic profiles who will develop MetS. We first performed online searches for identifying the pool of variants that influence MetS and its risk factors. Sources of information were: the NCBI Gene Entrez for MetS (332 candidate genes) and MetS risk factors BMI, WAIST, HDLC, TG, INS, GLUC, SBP and DBP, including biomarkers FIB, CRP, PAI-1, IL-6, ICAM-1, ADIP and WBCC. In addition, candidates for MetS (30 genes) and its risk traits from large GWAS were identified. Three main MetS domains were considered: obesity, insulin resistance and dyslipidemia. A variant/ candidate gene was considered to have a contribution to MetS, if it affected at least 2 major domains (obesity (BMI & WAIST) + dyslipidemia (HDLC & TG); glucose intolerance (INS & GLUC) (INS & GLUC) + HDLC & TG); BMI, WAIST + INS & GLUC). If a variant affects any 2 or more domains, then the MetS score of an individual is represented as the sum of each trait beta coefficient * recoded individual’s genotype (0,1,2)^1 a variable weight. The weight for traits from main domains is 1.0, whereas for blood pressure and any additional inflammatory marker is 0.25. This creates a MetS index, which can classify individuals with high and low risk for MetS. The scoring system designed is compared with data mining classifying methods (logistic regression, partition trees, k-nearest neighbors, and radial basis machine) on candidate genotypes in the Family Heart Study time 1 and time 2 (~8 years apart).

**1710T**

Should we account for the random effect of relatedness when using Principal Component Analysis in GWAS? M. de Andrade^1, J.P. Soller^2 1) Div Biomed Statistics & Informatics, Mayo Clinic, Rochester, MN; 2) Department of Statistics, University of São Paulo, SP, Brazil.

Studies of human complex diseases and traits associated with candidate genes are potentially vulnerable to bias (confounding) due to population stratification and inbreeding, especially in admixture population. In genome-wide association studies (GWAS) the Principal Components (PCs) method provides a global ancestry value per subject, allowing corrections for population stratification. However, these coefficients are typically estimated assuming unrelated individuals and if family structure is present and is ignored, such sub-structure may induce artificial PCs. Extensions of the PCs method have been proposed by Konishi and Rao (1992) taking into account only sibship relatedness and by Oualkacha et al. (2012) which can be applied to general pedigrees and high dimensional data. In this work we apply such analysis for estimation of global individual ancestry but admitting PCs extracted from different variance components matrix estimators. For the application we use the GENOA sibship data consisting of European and African American subjects and the Baependi Heart Study consisting of 80 extended families collected from the highly admixture Brazilian population, both with SNPs data from Affymetrix 6.0 chip. All the implementation are done using R package.

**1711F**

Genome-wide association study of ventricular fibrillation in the setting of acute myocardial infarction. R. Pazoki^1,2, J.S.S.G. de Jong^1, M.E. Adamski^3, N. Bruinsma^1, C.R. Dekkers^1, L. Bierkman^1, A.M. Wilders^1, M.W. Tamm^1, C.R. Bezrukov^1. 1) Department of Clinical Epidemiology, Biostatistics & Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands; 2) Department of Clinical and Experimental Cardiology, Academic Medical Center, Amsterdam, The Netherlands; 3) Department of Cardiology, Catharina Hospital, Eindhoven, The Netherlands.

Sudden cardiac death from ventricular fibrillation (VF) during acute myocardial infarction (MI) is a leading cause of total and cardiovascular mortality. Although a heritable component plays a role in the determination of risk for VF in the setting of acute MI, the underlying genetic factors remain largely unknown. We previously started exploring the role of common genetic variants in modulation of VF risk by conducting a genome-wide association study (GWAS) in the Arrhythmia Genetics in the Netherlands (AGNES) case-control set which consists of individuals with a first acute MI, where cases suffered VF and controls not. The aim of the current study was to carry out GWAS in an extended set of AGNES patients (171 cases, 310 controls) to identify additional susceptibility loci. In total, 1458 patients with a first acute MI, 686 with VF and 767 without VF were included in the current study. The AGNES cases and controls were genotyped using the Human610-Quad and HumanOmni2.5 Illumina arrays. Quality control and principle component analysis were performed using GenABEL package in R. Non-typed single nucleotide polymorphisms (SNPs) were imputed using data from HapMap. Logistic regression models implemented in the ProbABEL program was used to assess differences in the distribution of genotypes among cases and controls. The locus we previously identified on chr21 (rs2824292) exceeded the genome-wide significance threshold for association among cases and controls. The locus we previously identified on chr21 (rs2824292) exceeded the genome-wide significance threshold for association among cases and controls. The locus we previously identified on chr21 (rs2824292) exceeded the genome-wide significance threshold for association among cases and controls.

**1712W**

Genetic determinants of plasma levels of vitamin D binding protein in U.S. black women. E.A. Ruiz-Narvaez^1,2, L. Rosenberg^1,2, S.A. Haddad^1,2, J.R. Palmer^1,2. 1) Slone Epidemiology Center, Boston University, Boston, MA; 2) Department of Epidemiology, Boston University School of Public Health, Boston, MA.

**Background:** Vitamin D plays a key role in the regulation of a variety of metabolic processes such as calcium and phosphate absorption, xenobiotic detoxification, cell proliferation and differentiation, and immunomodulation among others. Vitamin D binding protein (DBP), the major carrier of circulating vitamin D metabolites, determines the free fraction or biologically active component of vitamin D. Genetic polymorphisms may in part determine circulating levels of DBP. Methods: In order to identify genetic variants that may explain variation of DBP plasma concentration, we genotyped 23 single nucleotide polymorphisms (SNPs) in vitamin D-related genes in 486 women of the ongoing cohort study Black Women’s Health Study (BWHS) who had plasma measure of DBP. We also genotyped 30 ancestral informative markers (AIMs) previously shown to provide a valid estimate of percent African (vs. European) ancestry. We used general regression models with backward stepwise selection process to find the best set of SNPs that explain variation of DBP plasma levels. Results: A model that included the 23 vitamin D-related SNPs plus individual African ancestry explained 83% of the variation of DBP plasma concentration; five SNPs were able to explain 82% of the DBP plasma level variation. Most of the DBP variation was explained by SNP rs7041, which represents an Asp-to-Glu change in the position 432 of DBP. The SNP rs7041 explained 74% of DBP plasma concentration variation, and each minor allele was associated with an increase of 2350 nmol/L of DBP (p<0.0001). Conclusions: We identified 5 SNPs that explain most of the variation of DBP plasma levels in African American women. These polymorphisms, through their association with DBP concentration, might be predictors of diseases involving vitamin D deficiency.
1713T
Bayesian Polygenic Risk Prediction using Summary Statistics. B.J. Vilhjalmsson 1,2, J. Yang 1, S. Lindstrom 1, A. Gusev 2, S. Ripke 2, N. Psaltopoula 2,5,6,7, R. Do 2,2, E. Stahl 8, P. Pasanovic 2, S. Pollock 2, N. Zaitlen 10, H.-H. Won 3,4, S. Kathiresan 3,4, M.E. Goddard 17, N. Wray 1, P.L. De Jager 2,5,6,7, M. Daly 2,2, P.M. Visscher 3, P. Kratt 10, N. Patterson 2, A.L. Price 1,2. 1) Department of Epidemiology, Harvard School of Public Health, Boston; 2) Broad Institute, Cambridge, MA, USA; 3) The University of Queensland, Brisbane, Queensland, Australia; 4) Massachusetts General Hospital, MA, USA; 5) Department of Anatomy, Brigham & Women’s Hospital, Boston, MA, USA; 6) Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA; 7) Harvard Medical School, Boston, MA, USA; 8) Icahn School of Medicine at Mount Sinai, New York, NY USA; 9) UCLA, CA Los Angeles USA; 10) UCSF, San Francisco, CA USA; 11) The University of Melbourne, Parkville, Victoria, Australia.

As sample sizes grow, it is warranted.

95%CI 1.15-2.22, p=0.005). Conclusions and Impact: Our community-based replication was applied to test for significance of genotype by cancer diagnosis interaction. Generalized estimation equation adjusted for familial correlation was used to test whether genetic variation at the TERT-CLPTM1L gene locus, has been shown to relate to lung, brain, bladder, and gastrointestinal cancer risk was observed (p=0.0008). Those carrying the minor allele of rs402710, a 60% lower lung cancer risk was observed (p=0.0008). Those carrying the minor allele of rs2736100 at the TERT gene were shown to have a a late age of onset for dementia: 92 years for 50% affected in both Caucasians and the HAAS cohort. 523 V/L carriers have been shown to have a allele frequency of 64% in Caucasions and 69% in Asians for the 523 V/L genotype. 523 allele carriers share the same haplotypes for the onset of dementia, with the ‘523 genotype providing a resolution of approximately 2-10 years between the median age of onset for dementia where 50% of the carriers of each genotype are affected. This study provides comprehensive genotypes of 523 allele carrier status and the onset of dementia and supports the generalization of the biomarker risk assignment algorithm to the Japanese ethnicity. Further work is needed to generalize the relationship between ‘523 genotype and age of onset of dementia across multiple ethnicities.

1715W
Genetic analysis of TOMM40 and APOE for the onset of dementia in the Honolulu-Asia Aging Study. M.W. Lutz 1, D. Goldgaber 1, D.K. Bums 1, A.M. Saunders 1, L. White 1, D. Rodgers 1, S.M. Sultana 1, H. Chatterjee 1, D.G. Goldgaber 1. 1) Duke University Medical Center, Department of Neurology, Durham, NC; 2) Stony Brook University, Stony Brook, NY; 3) Zinfandel Pharmaceuticals Inc., Durham, NC; 4) Pacific Health Research and Education Institute, Honolulu, HI.

Understanding the impact of genetic factors on the onset of dementia and Alzheimer’s disease requires longitudinal studies with accurate determination of age of symptom and/or disease onset. Recent work by Rosas et al. showed the genetic association of variants in both APOE and TOMM40 with Alzheimer’s disease (AD) risk. Here we present the first longitudinal observational study of Alzheimer’s disease in a large, long-running, longitudinal, North American, Asian, and Pacific Islander population. Specifically, an informative intronic poly-T variant (rs10524923 or ‘523) in the TOMM40 gene was shown to provide more precise stratification of age of onset of Alzheimer’s disease than APOE 4 carriers. We used a Cox proportional hazards regression model to test whether genetic variation at the TERT-CLPTM1L gene locus, has been shown to relate to lung, brain, bladder, and gastrointestinal cancer risk was observed (p=0.0008). Those carrying the minor allele of rs402710, a 60% lower lung cancer risk was observed (p=0.0008). Those carrying the minor allele of rs2736100 at the TERT gene were shown to have a a late age of onset for dementia: 92 years for 50% affected in both Caucasians and the HAAS cohort. 523 V/L carriers have been shown to have a allele frequency of 64% in Caucasions and 69% in Asians for the 523 V/L genotype. 523 allele carriers share the same haplotypes for the onset of dementia, with the ‘523 genotype providing a resolution of approximately 2-10 years between the median age of onset for dementia where 50% of the carriers of each genotype are affected. This study provides comprehensive genotypes of 523 allele carrier status and the onset of dementia and supports the generalization of the biomarker risk assignment algorithm to the Japanese ethnicity. Further work is needed to generalize the relationship between ‘523 genotype and age of onset of dementia across multiple ethnicities.

1716T
Molecular screening of CFTR Gene mutations in North Indian Asth- matic Children: a case-control study. P. Dixit 1, S. Awasthi 2, N. Maurya 2, S. Agarwa 2, M. Srivinasan 1, 1) Dept. of Paediatrics, King George’s Medical University, Lucknow, India; 2) Dept. of Medical Genetics, S.G.P.I.G.M.S., Lucknow.

Background & Objective: Asthma is a complex genetic disorder. Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene is an asthma susceptibility gene and carriers for this gene may develop obstructive pulmonary disease like bronchial asthma. Therefore, the objective of the study was to assess the association of CFTR gene mutations in genetic susceptibility to bronchial asthma. Material and Methods: A hospital based case-control study was carried out to compare 250 bronchial asthma cases and 250 age and sex matched controls, aged 5 months-15 years were recruited for the study. Cases included were those children presenting symptoms of asthma and excluded were clinically suspected cystic fibrosis (CF) or sweat chloride level > 60 mmol/L or suffering from other respiratory diseases. Spirometry and sweat chloride test were performed in all subjects. All cases were further categorized into four different categories as per Global Initiative for Asthma criteria (GINA): mild intermittent (83), persistent (96), moderate persistent (52), and severe persistent (19). Genotypingwas performed for 24 CFTR gene mutations; R553X, N1303K, D1145G, ‘523 genotype providing a resolution of approximately 2-10 years between the median age of onset for dementia where 50% of the carriers of each genotype are affected. This study provides comprehensive genotypes of 523 allele carrier status and the onset of dementia and supports the generalization of the biomarker risk assignment algorithm to the Japanese ethnicity. Further work is needed to generalize the relationship between ‘523 genotype and age of onset of dementia across multiple ethnicities.

Posters: Statistical Genetics and Genetic Epidemiology
1717F The Association of HLA-DQB1 Alleles with Nonalcoholic Fatty Liver Disease in Turkish Population. S. Katirni1, L. Doganay1-2, Y. Colak1, E. Senanes1, O. Ozturk1, C. Ulusogu2, N. Karatas1, I. Tuncer1. 1) Molecular Biology and Genetics, Istanbul Technical University, Istanbul, Turkey; 2) Department of Gastroenterology, Medeniyet University, Goztepe Teaching and Research Hospital, Istanbul, Turkey; 3) Department of Gastroenterology, Umme Hafsa Teaching and Research Hospital, Istanbul, Turkey; 4) Department of Gastroenterology, Dicle University, Medical School, Diyarbakir, Istanbul.

Introduction: Nonalcoholic fatty liver disease (NAFLD) is probably the most common liver disease in western and westernized countries, with an overall prevalence up to 20%. The spectrum of NAFLD grades from simple fatty liver (hepatic steatosis) to nonalcoholic steatohepatitis (NASH) which is its more progressive form. Hepatocellular apoptosis caused by activation of immune cells is an important feature contributing to the pathogenesis of NAFLD. Highly polymorphic human leukocyte antigen (HLA) gene is closely associated with immune response. The role of HLA class II molecules is to present antigens to CD4+ T lymphocytes which are recently shown to be increased in NASH patients (Inzagharat et al., 2011). Moreover, recent studies emphasize that HLA-DQB1 polymorphisms may also have an effect on the progression of NAFLD and thus, we aim to investigate the association between HLA-DQB1 alleles and NAFLD in Turkish population. Methods: A study group of 93 biopsy proven NAFLD patients (46 male and 47 female, mean age 42±10.2 years) and a gender and age matched comparison group consisting 101 healthy people (48 male and 53 female, mean age 41±9.5) are recruited. The alleles of HLA-DQB1 are detected by sequence specific primer (SSP) PCR reaction-restriction fragment length polymorphism and DNA sequencing (Olerup et al., 1993). Results and Conclusions: Among thirteen HLA-DQB1 alleles analyzed in this study, DQB1*0604 is significantly more frequent among the NAFLD patients compared to healthy controls (12.9% vs 5.9%, P=0.003, OR: 2.7, 95% CI: 1.2-5.7). In addition the frequency of DQB1*0302 is significantly higher in the healthy control group than the NAFLD patients (24.8% vs 7.5%, P=0.001, OR: 0.2, 95% CI: 0.1-0.6). These results suggested DQB1*0604 might be associated with NAFLD and DQB1*0302 might be a protective allele against NAFLD.

1718W Genetic susceptibility of FCER2 gene variants with asthma and its severity in north Indian children: a case-control study. N. Sharma1, S. Awasthi1, S. R Phadke2. 1) Department of Paediatrics, King George's Medical University, Lucknow-UP, Lucknow, India; 2) Department of Medical Genetics, SGPGIMS, Lucknow-UP, India.

Background & objective: Several strong evidences are available for genomic factors predisposing asthma in children. Previous studies have shown that the activated low-affinity IgE receptor (encoded by FCER2 gene) plays a pivotal role in allergic immune response, regulation of immunoglobulin E (IgE) levels and other associated inflammatory mediators. Any disturbance in the delicately balanced immune response due to genetic variations in the form of single nucleotide polymorphisms (SNPs) with in FCER2 gene may result in insufficient functions of this immune modulator gene, which can enhance the susceptibility to asthma and its severity. The present study involves investigation of association of the SNPs rs2834072 and rs7249320 (located 7 bp 3’of exon 9 and intron respectively in FCER2 gene) with asthma and its severity. Methods: Case-control based genetic association study was performed among 550 children (275 asthmatic children and 275 normal controls) from north India (in-English population). For each SNP (rs34845087) of perilipins allows HSL to break down triglycerides in lipid droplets into glycerol and fatty acid chains. Several single nucleotide polymorphisms (SNPs) of PLIN1 have been shown to be associated with levels of lipid fractions. In this study the associations of three PLIN1 SNPs (rs2289487, rs894160, and rs1052700) and an HSL SNP (rs34845087) with total-cholesterol, HDL-cholesterol, LDL-cholesterol, and triglyceride levels were evaluated in 127 adult subjects living in Tonga. A multiple regression analysis adjusted for age, sex, and body mass index (BMI) revealed that a copy of rs894160-T allele significantly decreased serum LDL-cholesterol by 11.04 mg/dl (P-value = 0.0350) and also a copy of rs1052700-T allele significantly decreased serum LDL-cholesterol by 12.39 mg/dl (P-value = 0.0248). No other significant associations were observed. The rs894160 and rs1052700 were in linkage disequilibrium with each other (D’ = 0.87 and r2 = 0.73). The haplotype analysis revealed that a copy of risk haplotype (rs894160-T and rs1052700-T) was significantly decreased serum LDL-cholesterol by 12.33 mg/dl (P-value = 0.0193). Our results suggest that PLIN1 polymorphisms affect the levels of serum LDL-cholesterol in Tongan adults.

1719F Significant association of Perilipin polymorphisms with LDL-cholesterol level. I. Naka1, R. Kimura2, T. Inaoka1, Y. Matsumura4, J. Ohashi3. 1) University of Tsukuba, Tsukuba, Ibaraki; Japan; 2) Transdisciplinary Research Organization for Subtropics and Island Studies, University of the Ryukyu, Okinawa, Japan; 3) Department of Environmental Sociology, Faculty of Agriculture, Saga University, Saga, Japan; 4) Faculty of Health and Nutrition, Bunkyo University, Kanagawa, Japan.

Perilipins and hormone-sensitive lipase (HSL) play key roles in the cellular regulation of triglyceride deposition and mobilization. Perilipins, which are encoded by perilipin 1 gene (PLIN1) on 15q26, are primarily expressed at lipid droplet surface in adipocytes and steroidogenic cells. Phosphorylation of perilipins allows HSL to break down triglycerides in lipid droplets into glycerol and fatty acid chains. Several single nucleotide polymorphisms (SNPs) of PLIN1 have been shown to be associated with levels of lipid fractions. In this study the associations of three PLIN1 SNPs (rs2289487, rs894160, and rs1052700) and an HSL SNP (rs34845087) with total-cholesterol, HDL-cholesterol, LDL-cholesterol, and triglyceride levels were evaluated in 127 adult subjects living in Tonga. A multiple regression analysis adjusted for age, sex, and body mass index (BMI) revealed that a copy of rs894160-T allele significantly decreased serum LDL-cholesterol by 11.04 mg/dl (P-value = 0.0350) and also a copy of rs1052700-T allele significantly decreased serum LDL-cholesterol by 12.39 mg/dl (P-value = 0.0248). No other significant associations were observed. The rs894160 and rs1052700 were in linkage disequilibrium with each other (D’ = 0.87 and r2 = 0.73). The haplotype analysis revealed that a copy of risk haplotype (rs894160-T and rs1052700-T) was significantly decreased serum LDL-cholesterol by 12.33 mg/dl (P-value = 0.0193). Our results suggest that PLIN1 polymorphisms affect the levels of serum LDL-cholesterol in Tongan adults.

1720F Genetic risk models: model size and confidence intervals of the risk estimates. Y. Shan1, D.T. Smelser2, G. Tromp2, H. Kuvaniemi3, D.E. Weeks1. 1) Department of Biostatistics, Graduate School of Public health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 3) The Sigierved and Janet Weiss Center for Research, Geisinger Health System, Danville, PA, USA.

Disease risk estimation plays a very important role in disease prevention. With more precise risk estimates, people can make more informed decisions about disease prevention and treatment. Many studies, which estimate disease risk using logistic regression models applied to a given set of risk SNPs, have found that the larger the number of risk SNPs in the risk model, the greater the ability to predict the risk. However, the width of the confidence interval (CI) is often not considered in the evaluation of the risk model. In fact, when risk estimates are based on meta analysis results, confidence intervals are often not even estimated. Furthermore, a risk estimate with a larger CI from a larger model may not be practically better than a similar risk estimate with a smaller CI from a smaller model. Here, we explore the relationship of predicted disease risk, the size of the model, and the width of the CI of the risk using both simulated data and real data. We developed a formula and conducted a simulation of how much a person’s risk could change when a single SNP was added into the model. We found that the width of the CI depended strongly on the magnitude of the estimated risk. In addition, we found that the CI width appeared to be positively correlated with the size of the model and the width of CI. This work was supported by a Career Development Award on Translational Genomics, part of The Common-wealth Universal Research Enhancement (CURE) program of the Pennsylvan-ia Department of Health.
1721W Genetic Risk Score Modeling in Age-Related Macular Degeneration.
J.N. Cooke Bailey1, J.D. Hoffman1, L.M. Olson1, W. Cade1, N. Schnitzlein-Boutaud1, P. Mayo1, M. Allen1, A. Agarwala1, M.A. Brantley1, W.K. Scott1, M.A. Penicak-Vance1, J.L. Haines1. 1) Center for Human Genomics Research, Vanderbilt University Medical Center, Nashville, TN; 2) Hussman Institute of Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 3) Vanderbilt Eye Institute, Vanderbilt University Medical Center, Nashville, TN.
Age-related macular degeneration (AMD) is a progressive neurodegenerative disease that is the leading cause of blindness in elderly individuals in developed countries. AMD risk is mediated by genetic and environmental factors. The majority of identified genetic risk for AMD appears to be influenced by single nucleotide polymorphisms (SNPs) in ARMS2/HTRA1 and CFH genes, which account for approximately 50% of the genetic risk for AMD. Along with several additional loci of lower size effect accounting for less than 65% of the genetic component of AMD. Most recent publications implicate 19 SNPs as the known mediators of AMD risk (Fritsche et al, Nat Gen 2013, 45: 433). In a Caucasian sample of 1196 AMD cases (individuals with intermediate and/or advanced AMD), we detected a significant difference in cumulative risk scores between cases and controls. Interestingly, when patients with exudative AMD were compared to those with bilateral exudative AMD, the risk score trended toward being higher in the fellow eye. To those with exudative AMD patients (AREDS grade 2-4 in the AREDS trial), the MetS prevalence is estimated to be 27.5% with regional variations, being highest in the Alentejo (30.99%) and lowest in the Algarve (24.42%), constituting a public health problem. Although for clinical settings, a binary definition of MetS enabling a dichotomization of a continuous outcome variable reduces the statistical power of the MetS association studies. Therefore, the aim of the present study is to identify genetic risk factors involved in MetS etiology, using a continuous MetS score. To achieve our goal, a principal component analysis was performed to compute a score using the six normalized risk factors for MetS (waist circumference, diastolic and systolic blood pressure, glucose, triglycerides and HDL blood levels), with a higher MetS score indicating a less favorable MetS profile. After calculating this score, an association analysis was performed using 37 SNPs in candidate genes involved in MetS related diseases. A total of 206 subjects, including 119 women and 87 men (mean age: 56.31±16.37 years, range: 26-91 years) were included in this analysis. We found 4 SNPs significantly associated with higher MetS scores (rs4244265 (CYP2C19), rs2735711 (GA- CYP2C19), rs4243454 (CA-MET), rs4244285 (CYP2C19) and rs4244285 (CYP2C19) and MetS) which shows the importance of lipid metabolism, thought cytochrome P450 enzymes, in the MetS etiology. However, further studies will be necessary to replicate these findings in different populations and to conduct functional studies to clarify the role of this variant in the etiology of MetS.

1722T Genetic variation at the CY2C19 gene associated with Metabolic Syndrome susceptibility in a South Portuguese population. V. Gaião1, A. Fernandes2, F. Menendez3, F. Cornejo Correia4, A. Beloza5, A. Gil5, M. Bourbon5, A. Vicente5, C. Dias5, M. Barreto da Silva5. 1) Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal; 2) Laboratório de Saúde Pública Dra. Laura Ayres, Faro, Portugal; 3) Administração Regional de Saúde do Algarve, Faro, Portugal.
Metabolic syndrome (MetS) is a cluster of conditions – increased blood pressure, high blood glucose level, excess body fat around the waist and abnormal cholesterol levels – that occur together, increasing the risk of heart disease, stroke and diabetes. In Portugal, the MetS prevalence is estimated to be 27.5% with regional variations, being highest in the Alentejo (30.99%) and lowest in the Algarve (24.42%), constituting a public health problem. Although for clinical settings, a binary definition of MetS enabling a dichotomization of a continuous outcome variable reduces the statistical power of the MetS association studies. Therefore, the aim of the present study is to identify genetic risk factors involved in MetS etiology, using a continuous MetS score. To achieve our goal, a principal component analysis was performed to compute a score using the six normalized risk factors for MetS (waist circumference, diastolic and systolic blood pressure, glucose, triglycerides and HDL blood levels), with a higher MetS score indicating a less favorable MetS profile. After calculating this score, an association analysis was performed using 37 SNPs in candidate genes involved in MetS related diseases. A total of 206 subjects, including 119 women and 87 men (mean age: 56.31±16.37 years, range: 26-91 years) were included in this analysis. We found 4 SNPs significantly associated with higher MetS scores (rs4244265 (CYP2C19), rs2735711 (GA- CYP2C19), rs4243454 (CA-MET), rs4244285 (CYP2C19) and MetS) which shows the importance of lipid metabolism, thought cytochrome P450 enzymes, in the MetS etiology. However, further studies will be necessary to replicate these findings in different populations and to conduct functional studies to clarify the role of this variant in the etiology of MetS.

1723F Hearing function and loss: a complex multistep strategy to identify genes and environmental factors. G. Girotto1, D. Vucković1, A. Buniel-lo1, K. Steel1, P. Gasparini1. 1) Medical Genetics IRCCS ‘Burlo Garofolo’-DMS, Univ.Trieste, Trieste, Italy; 2) King’s College London, Guy’s Campus, London, United Kingdom.

The analysis of complex genetic traits/diseases such as normal hearing function (NHF) has long been an enigma of genetic biology and medical sciences and the majority of genes/loci and environmental/lifestyle factors involved still need to be detected. Here we propose a multistep approach based on: A) Genome-Wide association study (GWAS) to detect association between candidate proteins and MetS, B) replication with general association studies and D) pathway analysis and E) genotype-phenotype relationship to detected genetic factors. As regards environmental factors a linear/logistic regression has been performed. For the genetic studies, we have selected a list of 19 genes recently identified in a large GWAS on different quantitative hearing traits (step A) carried out on 3417 individuals from isolated populations/localities located in Europe. These genes have been evaluated at the expression level, 12 of them showing a clear staining in the mouse cochlea of wild-type mice (at 4 and 5 days postnatal). In particular, 5 of them (Arsgr, Slc16a6, Dclk1, Gabrb3, Csdm1) show strikingly specific expression in the cochlea (e.g. of the top of sensory hair cells and in the marginal cells of the stria vascularis) whereas the other 7 (Pptrd, Grn, Gm2agb1, Evils, RmpB2, Ank2, Cdh13) are located in multiple cell types in the cochlea (step B). In the next step (C), 9 out of these 12 genes have been successfully replicated in independent cohorts from Caucasus and Central Asia. Moreover, 8 of these replicated genes fit in the same pathway which include a known HL gene (MARVELD2) (step D); finally, to look for genotype-phenotype relationship, the audiometric profiles (i.e. mean values at each frequency) of the 3 genotypes of the associated gene-variant have been analyzed. In particular, 7 genes out of the 9 replicated (CDH13, GM2AGB1, ANK2, SLC16A6, ARSG, RIMBP2 and DCLK1) showed a peculiar audiometric pattern ware a relevant differences for each genotype further supporting their role in NHF (step E). As regards environmental factors, among all the investigated variables (smoking, chocolate, coffee, tea, wine, beer, dairy products, spirits), only coffee consumption was associated at low and high frequencies (p=0.006) while the intake only at high frequency (p=0.003). These data demonstrate the usefulness of these combined methods in providing new insights into the molecular basis of NHF and may suggest new targets for hearing impairment treatment and prevention.
GWAS on an admixed Chilean sample of Cases and Controls to identify the genetic basis of the phenotypic variability in 22q11 microdeletion syndrome. S. Eyheramendy1, F. Maney1, M. Ramirez1, C. Vial2, K. Espinoza2, J.C. Rivera2, G. Repetto1. 1) Statistics, Pontificia Univ Catolica, Santiago, Chile; 2) Center for human genetics, Faculty of Medicine, Clínica Alemana, Universidad del desarrollo.

In this study we are interested in finding the genetic basis of the incomplete penetrance of the cardiovascular phenotype in 22q11 microdeletion syndrome (22q11DS) in a Chilean admixed sample of cases and controls. Genome-wide association studies (GWAS) have been a popular strategy to try to localize the genes or mutations that increase/decrease disease susceptibility in homogeneous populations. Unfortunately, the standard methodology utilized in GWAS needs to be adapted in the study of admixed populations. In order to assess for the association of a genetic variation with a disease, it is necessary to know the disease status of the individuals as well as the local ancestry of the mutations on the individuals. Chilean history shows that the general Chilean is mainly a mixture between Europeans (predominantly people from Spain but also from Germany, France, UK, Croatia, etc.) and Native Americans, in contrast with other Latin American populations which have an African component as well. In this study we performed a comparison of local ancestry estimation methods in terms of their performance and sensitivity towards different choices of ancestral populations in the association outcome. We consider ancestral populations from Native American, Europe and Africa to perform the comparison and showed that in fact the association outcome can be sensitive to the choice of ancestral populations and local ancestry estimation algorithm. We also performed a genetic structural analysis to estimate distances between individuals of different ancestry. Funded by Fondecyt-Chile Grants 1120813, 1100131 y 1130392.

Multi-step LASSO approach identifies a chromosomal segment associated with working memory. V. Freytag1, L. Gschwind2, A. Minnik3, D. de Quervain2, 3, A. Papassotiropoulos1, 3, 4, C. Vogler1, 2, 4. 1) Molecular Neuroscience, University of Basel, Basel, Switzerland; 2) Psychiatric University Clinic, University of Basel, Basel, Switzerland; 3) Cognitive Neuroscience, University of Basel, Basel, Switzerland; 4) Department Biozentrum, Life Sciences Training Facility, University of Basel, Basel, Switzerland.

Background: Working memory is a heritable trait, likely to involve a large number of variants with small effect sizes as it has been reported for other complex traits, e.g. human height. Several studies have already applied GWAS, starting to elucidate the molecular underpinnings of human working memory. Yet, a large fraction of the genetic contribution remains to be observed phenotype. Resulting p-values were corrected for the number of multiple tests performed.

Methods: An initial sample of N=1039 healthy young Swiss individuals that underwent working memory assessment using the n-back paradigm was subsequently subjected to genotyping with the Affymetrix SNP Array 6.0. We followed a multi-step approach, starting with pruning SNPs for strong LD (r²>0.8) on a genome-wide level. SNPs showing an uncorrected nominal (p<0.05) association signal were partitioned into chromosomal segments of 20 MB pairs. Within each segment a penalized regression model (LASSO) was trained on the remaining SNPs and combined with a bootstrap-based stability selection procedure to identify subsets of informative SNPs. Statistical significance of the subsets of SNPs was assessed in a second independent replication sample of N=719, by testing the correlation (unilateral Pearson’s correlation test) between the model prediction and the actual observed phenotype. Resulting p-values were corrected for the number of multiple tests performed.

Results: Several subsets of SNPs reached nominal significance. On chromosome 5q34-35, a set of 46 markers exceeding a stability fraction of 0.9 survived correction for multiple comparisons (uncorr. p < 0.00003, corr. p < 0.007), suggesting involvement of this region in working memory. Among those, 19 SNPs mapped to intragenic regions, yielding 10 candidate genes for further investigation.
1728T
Localization of causal variants at loci with multiple signals of association. F. Hormozdiari1, E. Kostem1, E. Kang1, B. Pasanovic1,2, E. Eskin1,2,4
1) Computer Science, UCLA, Los Angeles, CA; 2) Department of Human Genetics, UCLA, Los Angeles, CA; 3) Department of Pathology and Laboratory Medicine, Geffen School of Medicine, UCLA, Los Angeles, CA; 4) Inter-Departmental Program in Bioinformatics, UCLA, Los Angeles, CA.

Although genome-wide association studies have successfully identified thousands of risk loci for complex traits, only a handful of causal variants have been successfully identified. Several fine-mapping studies are currently underway in an effort to pinpoint causal variants for common disease. Although several loci have been shown to contain multiple causal variants, current statistical methods for fine-mapping rely on simplistic assumptions of a single causal variant per locus to estimate posterior probability of variants being causal (WTCCC et al Nature Genetics 2012). In this work we propose a new framework for statistical fine-mapping for causal variants. As opposed to current approaches that make assumptions about the number of causal variants at a given locus, our framework considers all models with arbitrary number of causal variants to properly estimate the posterior probability of a SNP being causal. A direct benefit of our approach is that we provide well-calibrated confidence sets of SNPs that are guaranteed to contain the true causal SNPs with high confidence (i.e. 95% confidence). In other words not all the SNPs picked in our set are causal but we claim we capture the true causal SNPs with high probability (probability higher than 95%). We assessed the performance of our method using both simulated and real data. We indicate in this work the idea of peaking the top significant SNPs or using the conditional method, which are the common methods to find the causal SNP, have high false negative rate. Using the simulation data we can illustrate our methods obtain a set of SNPs which contain the true causal SNPs with 95% probability and the average number of SNPs picked by our method is only twice the number of true causal SNPs implanted in the region.

1729F
A Phenome-Wide Association Study (PheWAS) Exploration of Multiple Traits at Baseline in AIDS Clinical Trial Group (ACTG) Protocols. C.B. Moore1,2, A. Verma3, D.H. Johnson1, E.S. Daar1, R.M. Gilikci4, R. Haubrich5, G.K. Robbins6, S.A. Pendergrass7, D. Haas1, M. Ritchie1, 1) Vanderbilt University, Nashville, TN; 2) Center for System Genomics, The Pennsylvania State University, University Park, PA; 3) Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 4) Weill Cornell Medical College, New York, NY; 5) University California San Diego, CA; 6) Harvard University, Cambridge, MA

Phenome-wide association studies (PheWAS) evaluate the associations between genetic variation and wide range of complex traits. We applied this approach to the baseline data from four AIDS Clinical Trial Group (ACTG) protocols using > 30 pre-treatment laboratory measures from 2547 individuals of mixed ancestry. We imputed data using Impute2 and a cosmopolitan reference panel, resulting in 5,954,294 SNPs. Due to the unique nature of this cohort, we divided the data to seek replication of our PheWAS results. We used our framework to calculate associations for each laboratory measure, adjusting for top principle components, age, and sex. To assess replication we compared results from study A059s to the results from combined studies A384, A5142, and A5202. A total of 1319 SNPs had replicating significance (p-value ≤1e-03, replicated p-value ≤1e-01). Top results were associated with total bilirubin and mapped to UDP-glucuronosyltransferase (UGT), responsible for glucuronidation (top result: test p-value 5.44e-35, replicated p-value 7.69e-30). Members of the UGT family have been previously associated with Gilbert syndrome and hyperbilirubinemia. The second cluster of SNPs were associated with high-density lipoprotein (HDL), located on chromosome 16 near or in CETP (top result: test p-value 5.60e-07, replicated p-value 3.65e-07). These SNPs have previously associated with serum cholesterol and metabolic measures. A total of 49 SNPs were associated with more than one measurement. For example, 35 SNPs were found to be significant for low-density lipoprotein measurements (LDL) and total cholesterol. These SNPs mapped to COPG2, TSGA13, and FAR2. The coatomer gamma subunit (COPG2) is a component of the protein complex required for budding from Golgi membranes. It influences the Golgi structural integrity and processing, activity, and endocytic recycling of LDL receptors. TSGA13 is in close proximity to COPG2. FAR2 encodes a dehydrogenase/reductase enzyme involved in converting fatty acids to fatty alcohols. In summary, we were able to identify robust SNP associations for multiple baseline laboratory measurements. Our results serve as a validation of whole-genome PheWAS, and show the utility of this approach for clinical studies. Future work includes PheWAS with additional ACTG data, including on-treatment data.

1730W
Fast And Robust Association Testing For High-Throughput Testing. Y.H. Zhou, F.A. Wright. Biostatistics, University of North Carolina, Chapel Hill, NC.

In the analysis of genotype-phenotype associations, the investigator often relies on parametric testing for high-frequency variants, but must resort to exact testing or permutation analysis for low-frequency variants. Such a two-stage procedure is cumbersome, and this difficulty is compounded for datasets with numerous rare variants. We describe the moment-corrected correlation (MCC) method for association testing, which provides a close approximation to exact test p-values, but with greatly reduced computation. The approach can be used for any phenotype distribution, and covariates can be handled by residualization or using covariate stratification. MCC is very fast, and can perform the association analysis for a genome scan in minutes. MCC is also, very general and it can be applied on others type of high dimensional ‘omics data, including gene expression data from microarrays or RNA-Seq. We illustrate the wide applicability of MCC as a screening and testing tool using a variety of ‘omics datasets.

1731T
A Novel General Framework for Imaging Genetics Analysis with Next-generation Sequencing Data. N. Lin, M. Chen, M. Xiong. University of Texas School of Public Health, Houston, TX.

Imaging Genetics is to investigate the relationship between the individual genetic variation and variation in brain wiring (connections), structure, and functional integrity. The traditional statistical methods for imaging genetics data analysis often simply summarizes three dimensional image data into an overall image measurement and tests association between a single genetic variant with the summarized simple statistic. These methods ignore image variation at positional level and use high dimension data reduction techniques to develop a statistic that collectively test association of all genetic variants within a gene or a genomic region with medical images. By intensive simulations, we demonstrate that the three dimensional functional linear model for imaging genetics analysis has the correct type I error rates and much higher power to detect association than the current methods. The proposed method was applied to glioblastoma (GBM) dataset where we extract MRIs of 78 GBM patients present in the Cancer Imaging Archive (TCGA) corresponding to patients in the Cancer Genome Atlas (TCGA) and RNA-seq and Exom sequencing data of 78 same GBM patients from TCGA dataset. Our preliminary results show that the proposed statistic has much smaller P-value than that of traditional method.
1732F
A Bayesian Approach to Expression Quantitative Trait Loci (eQTL) Mapping Based on Biological Pathway Knowledge. I.S. Chang1,2, T.Y. Chen3, C.H. Chen3, C.A. Hsiung1,2 1) National Institute of Cancer Research, National Health Research Institutes, Miaoli, Taiwan; 2) Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan; 3) Institutes of Statistics, National Tsing-Hua University, Hsinchu, Taiwan.

While genome-wide association studies (GWAS) have successfully discovered and replicated thousands of SNPs associated with various traits/diseases, it is a challenge to gain biological insights regarding this association, because over 80% of them are not in coding region. One of the main approaches in the follow up of a GWAS is to examine whether these SNPs from GWAS are associated with the expression levels of certain genes. This leads to the so-called expression quantitative trait loci (eQTL) studies. In the current approaches, it is common to consider the expression data of 25K probes and as many SNPs as possible simultaneously so as to explore new phenomena or generate new hypotheses. While intuitive, these approaches often ignore the correlation between the gene expressions and do not make use of biological pathway information. We propose a Bayesian approach to eQTL that alleviate these concerns. In particular, the prior distribution makes use of the classical heritability concept in genetics and the pathway information from biological databases like GO, KEGG, BIO-CARTA, etc. The former helps to avoid the often too conservative practices in genomic studies and the latter helps to provide biological interpretation and avoid reducibility issue. In addition, the true-path rule is considered in the model construction, which says that if a gene is associated to a pathway, it is also associated to all the pathways along the path up to the root in the tree. Finally, we carefully designed MCMC algorithm is proposed to introduce a substantial speed up of the permutation approach, wherein test statistic data from the Psychiatric GWAS Consortium (PGC) stage 1 and detected a number of interesting and unexpected interactions among SNPs in various pathways.

1733W
A Novel Exact Test for Association for Small Sample Case-Control Studies. L. Ehwerhemuepha, S. Alexandria, C. Rakovski. Chapman University, Orange, CA.

We propose a novel exact test for association between a multiallelic marker and a phenotype with small sample case-control data. In these settings, the case-control genotype data follow multinomial distributions but classical large-sample chi-square contingency table methods are not applicable. The approach enumerates all samples under the non-completely specified null hypothesis of equality of the underlying multinomial distributions and calculates p-values as the sum of the probabilities of the samples as likely or less likely to occur than the observed data. We performed an extensive simulation study to assess the type I error rates under various null hypothesis and at several alpha levels. The number of all possible samples is a product of large binomial coefficients that make the full calculation computationally intensive. Thus, we developed a fast version of the algorithm that reduces the computational time by a factor of 1000 based on the idea of a selective removal of samples with very low probabilities while controlling the precision of the estimated exact p-values. Our results show that the new method possesses a conservative type I error in all scenarios due to the absence of adjustment in this nonparametric technique for the estimation of common multinomial probabilities under the null. It is a viable association approach that attains moderate power to detect deleterious mutations with very large effect sizes even with small sample data. The proposed method is readily extendable to haplotype data and even multimarker genotype data with haplotypes phase uncertainty. In the latter case, the EM algorithm can be implemented to determine weighted haplotype assignments of all subjects based on the haplotype pairs compatible with the unphased genotypes.

1734T
SNPx: fast testing of SNPxSNP interactions using only summary statistic data. S. Baca, T. Bigdell, D. Lee. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA. Analyses of genome wide association studies (GWASs) and meta-analyses have mainly focused on detecting univariate association between trait and single nucleotide polymorphisms (SNPs). However, SNPxSNP interactions are likely to play important role in the etiology of complex phenotypes/diseases and, thus, detecting such interactions might provide new insights into the biological basis of these phenotypes. Nonetheless an exhaustive testing for these interactions has been prohibitively difficult due to i) the lack of access to individual level genotype data used in GWASs/meta-analyses and ii) extremely heavy computational burden required for a) the imputation of unmeasured SNPs and b) testing of all pairwise SNP interactions. To alleviate the burden of testing SNPxSNP interactions, we propose a novel method/software for detecting interaction between SNPs in a genome scan (GWAS or whole genome sequence scan), by using only summary statistics. The proposed method employs the conditional expectation formula for multivariate normal variates to impute i) summary statistics of unmeasured SNPs and ii) interaction statistics between all nominally significant SNPs. Based on these univariate summary statistics and their correlations, as estimated from a relevant reference panel (e.g. 1000 Genomes and UK10k), we construct the omnibus SNPxSNP test. We applied the method to summary statistic data from the Psychiatric GWAS Consortium (PGC) stage 1 and detected a number of interesting and unexpected interactions among SNPs in various pathways.

1735F

Set tests are rapidly gaining in importance, proving useful for testing associations between phenotypes and sets of (possibly rare) variants defined by genes, pathways, or regions of the genome. We introduce a new methodology for such tests that yields improved power over widely-used kernel-based methods (such as SKAT), while controlling for type-I error. The test we consider uses a test statistic that is the difference in out-of-sample (cross-validated) phenotype prediction accuracy between using all and no variants in the set. This statistic has been used previously, but obtaining P values has been computationally inefficient for genome-wide testing, because it has required a huge number of permutations for reliably estimating the extreme tail of the null distribution, the region of most interest. Here, we introduce a substantial speed up of the permutation approach, wherein test statistics for a small number of permutations are computed (typically, 10 permutations per test), and then the tail of the distribution of pooled test statistics is fit to a simple parametric null distribution. One example of a parametric distribution that we have found to work well in practice is a mixture distribution between a constant zero component and a gamma distribution. We illustrate the approach on several synthetic and real data-sets, including data from the Wellcome Trust Case Control Consortium. The approach is extremely general as it can be used with various linear and non-linear classification/regression models and various measures of prediction accuracy and does not rely on asymptotic assumptions.
1736W Analyzing genome-wide associations with high dimensional phenotypes in the GALA II study. C.R. Gignoux1, J.M. Galanter1, K.A. Drake1, H. Aschard2, D.G.Torgerson1, A.L. Roth3, S.S. Oh1, P. Kraft4, C.D. Bustamante3, N.A. Zaitlen1, E.G. Burchard1, The GALA II Investigators. 1) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 2) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 3) Department of Genetics, Stanford University, Stanford, CA.

Studies of complex traits often measure a large number of phenotypic variables beyond the specific measurement of interest. Ideally these can help to elucidate disease severity and identify potential disease heterogeneity. However, these measurements often contain missing data and can be highly correlated, noisy and difficult to interpret on their own. To identify genetic associations with high dimensional phenotypes, we leveraged an extensive number of measurements in the GALA II study of childhood asthma in Latinos.

We used a modified principal component analysis (PCA) capable of handling missing data to decompose the 75 phenotypes into uncorrelated components and performed a genome-wide association study for each PC using standard linear regression. We then performed a meta-analysis across all PC-GWASes to identify variants associated with any phenotype variables.

We found that for most PCs, scores were drawn from continuous distributions. Importantly the structure was not defined by a small number of PCs: PC1 explained ~8% of the total variance, while the top 10 PCs together still explained less than 40%. Yet the orthogonality ensured that PCA separated different components of the disease process: the top three PCs related largely to severity of disease (including frequency of symptoms and exacerbations), while PC7 summed downstream of GIPC2, previously believed to be CD specific, was shown to be shared. While the UC association alone was not significant (OR=1.05, p=0.07, compared to OR=1.13, p<10–5 in CD), the joint analysis showed strong evidence of shared risk (BF=100). This suggests a differential enrichment of functional terms (GO and KEGG) in CD and UC. Finally, we fine-map 78 densely genotyped loci by simultaneously considering data from both CD and UC, and compare the results to standard fine-mapping techniques.


There is growing interest in understanding the shared and distinct genetic risk pathways that underlie different human diseases. However, these investigations can be confounded by differential power and control sharing, making loci falsely appear distinct and shared respectively. To address these issues we need a cross-phenotype model that is capable of combining evidence across different diseases from multiple data sets. The method allows regression techniques such as covariates, conditional interaction, and testing specific hypotheses about locus sharing (e.g. showing a critical role in the myo-inositol biosynthesis pathway by catalyzing the rate-limiting conversion of glucose 6-phosphate to myo-inositol 1-phosphate). Our hypothesis is that SNP of SLC5A11 and ISYNA1 gene may affect the maternal MI concentration and then disturb the process of neural tube closure.

In the present study, we carried out a case-control study in a Chinese population of Shanki Province, a high-risk area for NTDs. The case group was consisted of 150 pregnant women with NTDs and there were 279 pregnant women with normal newborns in the control group. Totally 8 tagSNPs of SLC5A11 and ISYNA1 gene were genotyped by the MassARRAY platform and analyzed. The result showed all the 8 SNPs in the present study were in Hardy-Weinberg equilibrium. In SLC5A11 gene, CC+CT genotype of rs274077 decreased risk for NTDs compared to those harboring the TT genotype (OR=0.64, 95%CI: 0.42-0.96, P = 0.029). After stratifying NTDs group into anencephalus, spina bifida and encephalocoele, AG+GG genotype of rs38052776 was significantly associated with NTDs (OR=2.60, 95%CI: 1.13-6.05, P = 0.027). In ISYNA1 gene, none of the SNP was associated with the NTDs risk. Then we predicted the binding capacity to the potential transcription factor of the different genotype of the above 2 SNP using the bioinformatics method. The rs2740777 is located in the conserved sequence of signal transducers and activators of transcription (STATX), and rs805778 is located in the conserved sequence of transcription factor c-Rel and NF-kappaB. Our result suggested the combined effect of SLC5A11 and ISYNA1 gene may affect the transcription factor binding ability, therefore, might regulate the expression and transcription of SLC5A11 gene to increase the risk of NTDs.


Neural tube defects (NTDs) are among the commonest and most severe disorders of the fetus and newborn. Maternal nutritional status is associated with the occurrence of NTDs. Myo-inositol (MI) is an essential brain constituent, and development can cause cranial NTDs in both rats and mice, and some human NTD pregnancies also have lower maternal MI concentrations than unaffected pregnancies. In the inositol metabolic pathway, Solute carrier family 5 member 11 (SLC5A11) is exclusively responsible for apical MI transport and absorption in intestine and the encoded protein of Myo-Inositol 1-phosphate synthase A1 (ISYNA1) plays a critical role in the myo-inositol biosynthesis pathway by catalyzing the rate-limiting conversion of glucose 6-phosphate to myo-inositol 1-phosphate. Our hypothesis is that SNP of SLC5A11 and ISYNA1 gene may affect the maternal MI concentration and then disturb the process of neural tube closure.

In the present study, we performed a meta-analysis across all PC-GWASes to identify variants associated with any phenotype variables. Totally 8 tagSNPs of SLC5A11 and ISYNA1 gene were genotyped by the MassARRAY platform and analyzed. The result showed all the 8 SNPs in the present study were in Hardy-Weinberg equilibrium. In SLC5A11 gene, CC+CT genotype of rs274077 decreased risk for NTDs compared to those harboring the TT genotype (OR=0.64, 95%CI: 0.42-0.96, P = 0.029). After stratifying NTDs group into anencephalus, spina bifida and encephalocoele, AG+GG genotype of rs38052776 was significantly associated with NTDs (OR=2.60, 95%CI: 1.13-6.05, P = 0.027). In ISYNA1 gene, none of the SNP was associated with the NTDs risk. Then we predicted the binding capacity to the potential transcription factor of the different genotype of the above 2 SNP using the bioinformatics method. The rs2740777 is located in the conserved sequence of signal transducers and activators of transcription (STATX), and rs805778 is located in the conserved sequence of transcription factor c-Rel and NF-kappaB. Our result suggested the combined effect of SLC5A11 and ISYNA1 gene may affect the transcription factor binding ability, therefore, might regulate the expression and transcription of SLC5A11 gene to increase the risk of NTDs.

1739W Susceptibility loci do not infer cognitive impairment as measured by TICS-M in multiple sclerosis patients. M.F. George1, E. Elboudwarej1, F.B.S. Briggs1, H. Quach1, R. Whitmer1, L. Shen2, A. Bernstein3, C. Schaefer4, L.F. Barcellos1,2,1) Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, University of California, Berkeley, CA, USA; 2) Kaiser Permanente Division of Research, Oakland, CA, USA; 3) Palm Drive Hospital, Petaluma, CA, USA.

The prevalence of cognitive impairment may be as high as 70% in individuals with multiple sclerosis (MS) and has been shown to affect social and emotional function, maintenance of employment, and overall quality of life. The identification of genetic variants that predict severe clinical MS outcomes is critical to understand disease mechanisms and guide development of effective therapeutics. We investigated the association between established risk loci and disease progression measured by cognitive impairment. In total, 932 white MS patients and 576 controls identified from the Kaiser Permanente Medical Care Plan, Northern California Region were studied. A weighted genetic risk score (wGRS) combining the weighted odds ratio (OR) from each of 52 established MS risk loci was calculated for each allele counts per person. HLA-DRB1*15:01 carrier status was also examined. Cognitive status was determined using a validated telephone interview cognitive status (TICS-M) assessment tool, and accounted for level of education. Each participant was questioned regarding orientation; registration and free recall; attention and calculation; comprehension; language and repetition; and delayed recall. These seven areas of cognition were also examined as sub-scores. Linear regression models were utilized; adjusted models controlled for age, gender and known environmental risk factors. As expected, overall cognitive score was lower in MS cases compared to controls: p<10–6 in all seven areas of cognition. Sub-scores (orientation p=3×10–5; registration p=0.013, comprehension p=0.048, and delayed recall p=2×10–3). When testing in cases only, no association between individual risk variants, wGRS, GRs, or HLA-DRB1*15:01 and cognitive impairment was observed. We hypothesized that the relationship between genetic risk scores (weighted and unweighted) and established MS risk loci on cognitive status in MS cases. Results suggest genetic factors contributing to cognitive impairment in MS are different from those that predispose to disease onset.
**1740T**

The role of rare genetic variants in host genetic control of anti-mycobacterial immunity. J. Manry1-2, A. Cobat1,2, E. Schurr1,2. 1) McGill International TB Centre, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 2) Departments of Medicine and Human Genetics, McGill University, Montreal, Quebec, Canada.

The tuberculin skin test (TST) measures the intensity of anti-mycobacterial acquired immunity and is used to diagnose latent infection (i.e. without clinical symptoms) with Mycobacterium tuberculosis. In a previous genome-wide linkage study, we identified two loci that have an impact on TST reactivity. The TST1 locus impacts on TST zero versus nonzero while TST2 impacts on the intensity of the developing immune response. Present experimental data do not support a strong effect of variants with a minor allele frequency (MAF) > 5% on both phenotypes leading to the conclusion that TST responses are under control of rare variants. Exome sequencing was performed on 96 individuals from the founder generation of the families that led to the detection of the TST1 and TST2 linkage peaks. A bioinformatic pipeline for variant discovery was setup, and 2461 variants were identified in genes located within the two linkage intervals (1305 in the TST1 locus and 1156 in the TST2 locus). To reduce the number of candidate variants, several filters were applied to identify those variants that are most likely to impact on the two phenotypes. In a first step, the exclusion of variants that corresponded to synonymous amino acid changes or that were not in Hardy-Weinberg equilibrium allowed us to reduce the list to 430 variants. Next, variants that had previously been genotyped in a high resolution scan of common variants of the linkage intervals were eliminated reducing the list to 351 variants. Interestingly, 239 of these 351 variants were found in families contributing substantially to linkage (maximum LOD score for the family > 0.1). Indeed, 26 genes in the TST1 locus and 11 genes in the TST2 locus were only present in those families. This genotyping effort will allow us to identify genes with enrichment of variants that segregate with TST reactivities.

**1741F**

The Empirical Assessment of Statistical Power of Rare Variant Association Methods. K. Hao1,2,3, H. Chen1, H. Zhou1, C. Molony1, H. Dai1,2,4. 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Asian Cancer Research Group, Inc., Wilmington, DE; 4) Merck Research Laboratories, Boston, MA.

Background: The role of rare genetic variation in the etiology of complex disease remains as an active research field. The next-generation sequencing technologies were powerful in identifying genetic variants responsible to Mendelian disorders, but had limited success in dissecting complex diseases. Several novel statistical methodologies have been recently proposed to assess the contribution of rare variation to complex disease etiology. So far, their statistical power was assessed mainly on simulated phenotypes. Methods: We quantified the statistical power of popular rare variants association methods empirically using the eQTL framework. That is the relative statistical power is proportional to the number of eQTLs captured at fixed FDR. Results: On N=100 human non-tumor liver tissues, we measured gene expression trait with Affymetrix Hu133 Plus2 chip, and conducted whole genome DNA sequencing (WGS) at 30x coverage. Also, we measured the genotype using illumina 650Y array. Results of conventional GWAS methods with 1000G imputation was compared to those of rare variants association methods, including cohort allelic sum test (CAST), weighted sum test (WST) and SKAT. When only focusing on SNV of MAF<5%, none of the rare variants association methods have meaningful statistical power. When including all SNVs, rare variants association methods offer comparable power to single marker test (ie, GWAS methods) Conclusion: GWAS has no advantage in term of statistical power over SNP array in identifying eQTLs. Gene-level tests offer higher power due to less multiple testing penalties. SNVs on exons, splicing site or UTRs were more likely to control gene expression levels.

**1742W**

Enhanced ability to replicate findings as a rationale for conducting marker-set tests. J. Arbet1, K. Grinde2, C. Fu3, A. Benitez4, M. O’Connell5, N. Tinto6. 1) Winona State University, Winona, MN; 2) St. Olaf College, Northfield, MN; 3) Massachusetts Institute of Technology, Cambridge, MA; 4) Brown University, Providence, RI; 5) Miami University of Ohio, Oxford, OH; 6) Dordt College, Sioux Center, IA.

Over the last decade numerous approaches have been proposed to simultaneously analyze multiple individual variants for association with a phenotype. These approaches span methods which combine rare variants (minor allele frequency less than 5%), common variants, rare and common variants in pathway based approaches. All of these methods hope to provide improved power versus single variant approaches by (a) reducing multiple testing penalties and (b) aggregating evidence from multiple causal variants by creating biologically meaningful sets. While simulation studies have shown that for particular genetic architecture and based on the proportion of non-causal variants, multi-marker approaches can provide improved power in a discovery (stage 1) analysis, little work has explored the impact of single vs. multi-marker tests on the power to replicate findings in an independent (stage 2) study. We present general results for power in independent samples when using multiple marker tests vs. single marker tests, and demonstrate that the advantages of multiple marker approaches are a function of the number of markers combined, the proportion of non-causal variants, the strength and direction of the causal variant association with the phenotype and the multiple testing penalty being used. Practical guidance on use of multiple marker tests with a goal of maximal power for replication is provided.

**1743T**

Maximizing the power in Principal Components Analysis of Correlated Phenotypes. H. Aschard1, B. Vilhjalmsson2,3, C. Wu1, N. Greliche4, P.E. Morange5, B. Wolpin6, D. Tregouet2,7, P. Kraft2. 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) INSERM UMR S 937, ICAN Institute for Cardiometabolism And Nutrition, Pierre et Marie Curie University, Paris 6, France; 3) INSERM UMR_S 1062, Aix-Marseille Université, Marseille F-13385, France; 4) Dana-Farber Cancer Institute, department of Medical Oncology, Boston, USA.

Background: Principal Component analysis (PCA) is a useful tool that has been widely used for the multivariate analysis of correlated variables. It is usually applied as a dimension reduction method: the few top principal components (PCs) explaining most of total variance are tested for association with a predictor of interest, and the remaining PCs are ignored. This strategy has been widely applied in genetic epidemiology, however some aspects of this analytical technique are not well appreciated in the context of single nucleotide polymorphisms (SNPs) testing.

Method: We reviewed some of the theoretical basis and behavior of PCA when testing for association between a SNP and two correlated traits under various scenarios. We then evaluated through simulations the power of a few different PCA-based strategies when analyzing up to 100 traits. We then compared these methods for the genome-wide association test of five coagulation traits in 685 subjects from the MARTHA study, and the association test of 700 candidate SNPs with 79 metabolite levels in 1190 individuals from four U.S. prospective cohorts.

Result: We show that contrary to widespread practice, testing the top PCs only can be dramatically underpowered since PCs explaining a low amount of the total phenotypic variance can harbor a substantial part of the total genetic association. We also demonstrate that PCA-based strategies can only achieve a moderate gain in power in the presence of positive pleiotropy, but have great potential to detect negative pleiotropy (e.g. positive correlation and opposite genetic effects) or genetic variants that are associated with a single trait highly correlated to others. Real data applications confirm these results: the combined analysis of the five PCs from the coagulation traits identified two new potential candidates SNPs, which had strongest associations with the 5th PC, while the combined analysis of half of the PCs from the metabolite traits explaining the less of the total variance (12% of the total when combined) identified two new variants that likely affect a single trait.

Conclusion: We identified major improvements to standard PCA-based strategies for the analysis of correlated traits. Their implementations led to the identification of new genetic variants that would have been missed by standard approaches.
What now? Post-hoc approaches for gene-based, rare variant tests of association. A. Benitez1, J. Arbet2, K. Grinde3, C. Fu4, M. O’Connell5, N. Tintle6, 1) Brown University, Providence, RI; 2) Winona State University, Winona, MN; 3) St. Olaf College, Northfield, MN; 4) Massachusetts Institute of Technology, Cambridge, MA; 5) Miami University of Ohio, Oxford, OH; 6) Dordt College, Sioux Center, IA.

To date, rare variant testing approaches have focused on maximizing statistical power to identify genes showing significant association with disease. The statistics can accommodate a combination of risk, protective, and non-causal variants and can weight each variant. Although increasingly complex test statistics and weighting strategies may improve power, they also may make interpretation of a significant association result more difficult. Determining which variant(s) in the gene are causal and estimating their effect is crucial toward planning replication studies and characterizing the genetic architecture of the locus. This problem is analogous to the situation in a one-way ANOVA analysis, where evidence in the omnibus test says that ‘at least one group mean is different than the others,’ but it takes post-hoc analyses to identify specific group means as being different than others. Recent work by our group has classified general characteristics of two large classes of gene-based rare variant tests. Using this framework, we have explored the ramifications of choice of gene-based test statistic on post-hoc analyses attempting to identify causal variants. Furthermore, we have evaluated the overall quality and consistency of different single marker association statistics (e.g. single marker test p-values, difference in allele counts between cases and controls, relative risk estimates) in identifying the most likely causal variants within a gene. Finally, we have explored approaches to identify the most likely subset of causal variants within a gene, when incorporating a priori biological information (e.g., position, function). In this presentation we will present the results of our findings, ultimately providing a set of best practices for applied researchers.

1746T 8q24 Risk Alleles and Prostate Cancer in African-Barbadian Men. C.D. Cropp1, C.M. Robbins2, A.J.M. Haniss3, J.D. Carpten2, L. Waterman2, R. Worrell2, J.M. Trent2, M.C. Lesko4, S.Y. Wu5, J.E. Bailey-Wilson5, B. Nemeshure2. 1) Statistical Gen Branch, IDRB/NHGRI/NIH, Baltimore, MD; 2) Integrated Cancer Genomics Division, Translational Genomics Research Institute (TGen), Phoenix, AZ; 3) Department of Biological & Chemical Sciences, University of the West Indies, Bridgetown, Barbados; 4) Department of Preventive Medicine, Stony Brook University Medical Center, Stony Brook, NY.

African American men (AA) exhibit a disproportionate share of prostate cancer (PC) incidence, morbidity and mortality compared to other groups. Several genetic association studies have implicated select loci in the 8q24 region as increasing PC risk in AA. We evaluated the association between previously reported 8q24 risk alleles and PC in African-Barbadian (AB) men, also known to have high rates of PC. Ten previously reported tag SNPs were genotyped in 447 AB men with PC and 385 AB controls from the Prostate Cancer in a Black Population (PCBP) study. Only rs2124036 was nominally significant in AB men, (OR = 2.0, 95% CI (1.0-4.3), P=0.06) for the homozygous C/C genotype after correction for multiple testing. We also conducted a meta-analysis including our AB population along with two additional African-Caribbean populations from Tobago and Jamaica for SNPs rs16801979 and rs1447295. A significant association resulted for the rs16801979 A allele (Z score 2.75; p=0.006; summary OR= 1.21 (95% CI: 1.01-1.46)). Our findings may indicate: i) the presence of a founder effect; ii) the selected SNPs not being tagged to an ancestral haplotype bearing the 8q24 risk allele(s) in this population; or iii) inadequate power to detect a true association. Additional GWAS and sequencing studies are underway to further investigate any potential contribution of the 8q24 region to PC in this West African-derived population.

1747F Powerful methods for including genotype uncertainty in tests of Hardy-Weinberg Equilibrium. C. Fu1, M. O’Connell1, A. Benitez1, J. Arbet1, K. Grinde1, R. Liu1, A. Luedtke2, A. Beck3, N. Tintle4. 1) Massachusetts Institute of Technology, Cambridge, MA; 2) Miami University of Ohio, Oxford, OH; 3) Brown University, Providence, RI; 4) Winona State University, Winona, MN; 5) St. Olaf College, Northfield, MN; 6) Dordt College, Sioux Center, IA; 7) Harvard University, Cambridge, MA; 8) University of California- Berkeley, Berkeley, CA; 9) Loyola University Chicago, Chicago, IL.

Most genotyping technologies now generate posterior probabilities reflecting genotype uncertainty. Increasingly, methodological work seeks to utilize these posterior probabilities in downstream statistical analysis. In many cases, use of posterior probabilities (e.g., the dosage) can yield improved statistical power. Recent work by our group and others has shown that for standard single marker tests of association use of the dosage provides a nearly optimal use of posterior probabilities. However, prior to association testing, a standard quality control step in the analysis of SNP genotypes is to test for departures from Hardy-Weinberg Equilibrium (HWE). In this presentation we propose two novel strategies for testing HWE using posterior probabilities, and compare to a third recently proposed option. We demonstrate that these novel strategies provide improved power and proper control of the type I error rate as compared to existing methods.

1745W FARVAT: a fast and efficient rare variant association Test for dichotomous trait with extended families. S. Choi1, S. Lee1, M.M. Nöthen2, C. Lange3, T. Park1, S. Won4. 1) Bioinformatics Program, Seoul National University, Seoul, South Korea; 2) Institute of Human Genetics, University of Bonn, Germany; 3) Dept of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 4) Dept of Statistics, Seoul National University, Seoul, South Korea; 5) Dept of Applied Statistics, Chung-Ang UNiversity, Seoul, South Korea.

Family-based samples are genetically more homogeneous than the population-based samples and the analysis of rare variant with family-based samples have been often recommended. However in spite of the importance of family-based samples for rare variant association analysis, few statistical methods have been suggested. In this report we propose a new statistical method for the analysis of rare variant with extended families. The proposed method is robust against the population substructure, and can be applied to both binary and continuous traits. Depending on the choice of working matrix, our method can be a burden test or variance component test, and this approach is extended to the SKAT-O statistic. We empirically showed that the proposed methods perform better than the existing methods such as PedCMC and FPCA (MC Wu et al. AJHG, 2011). Furthermore the time complexity for the proposed method is O(M3 + N2M) and we found that the analysis of the whole genome sequence data for 1000 individuals in extended family design can be conducted within a few hours. The proposed method is implemented in C, and the proposed method was applied to gene-based analysis of Schizophrenia. All samples were collected from Germany, and the whole genomes for 36 trios which consist of affected offspring and unaffected parents were sequenced. There were 9,216,373 bi-allelic variants available, and 31,046 variants MAF of which minor allele frequencies were less than 0.05. Using 13,053 impact genes, we identified several candidate genes at the 5% significance level after Bonferroni adjustment, which are related to schizophrenia.
1748W

A clustering approach for mapping rare variants based on mutual association. S. Ghosh, S. Deb. Human Genetics Unit, Indian Statistical Institute, Kolkata, India.

Although genome-wide association studies have successfully identified a large number of common variants underlying various complex disorders, a substantial proportion of the total genetic variation in a trait still remains unexplained. It is becoming increasingly evident that the ‘Common Disease, Common Variant’ paradigm needs to be modified and the ‘missing heritability’ may possibly be explained by rare variants that could not be identified using genome-wide association studies. However, the major impediment in identifying rare variants is that one would require huge sample sizes to detect differences in allele frequencies between cases and controls. Thus, most existing methods are based collapsing multiple variant sites using different statistical algorithms. Motivated by the combined multivariate and collapsing (CMC) algorithm, we develop a clustering mechanism of rare variant sites, but based on their mutual extent of association rather than similarity in allele frequencies as proposed in CMC, thereby reducing the possibility of combining functional and non-functional variants. We use the Fisher’s exact test to identify blocks of variant sites such that the initial site in each block is associated with all other sites in the block. The test for case-control association is then performed within each block by comparing the proportions of affected and unaffected individuals carrying at least one copy of a rare variant and is based on a variance stabilizing sine transformation. We carry out extensive simulations under different rare variant models and compare the false positive rate and the power of our proposed method with some of the popular competing methods: CMC, adaptive SUM, WSS, TestRare, RareCover and Kernel-based Adaptive Clustering. We find that the proposed test procedure yields more power than the existing approaches, especially with increasing sample size, while maintaining the correct size.

1749T

Evaluating the impact of genotype errors and uncertainty on gene-based rare variant tests of association. K. Grinde1, C. Fu2, J. Arbet2, A. Benitez2, M. O’Connell2, N. Tintle2. 1) St. Olaf College, Northfield, MN; 2) Massachusetts Institute of Technology, Cambridge, MA; 3) Winona State University, Winona, MN; 4) Brown University, Providence, RI; 5) Miami University of Ohio, Oxford, OH; 6) Dordt College, Sioux Center, IA.

The new class of gene-based rare variant tests of association have usually been evaluated assuming error-free genotype information. In reality, rare variant genotypes, whether from next generation sequencing or imputation will reflect genotype uncertainty (e.g., genotype likelihoods, dosages or errors), and, ideally, subsequent rare variant tests should be robust to this uncertainty. Errors and uncertainty in SNP genotyping are already known to impact statistical power for single marker tests on common variants and, in some cases, inflate the type I error rate. Recent results show that uncertainty in genotype calls derived from sequencing reads are dependent on several factors, including read depth, calling algorithm, number of alleles present in the sample, and the frequency at which an allele segregates in the population. Imputation accuracy of rare variants is dependent upon the frequency of the allele to be imputed, as well as the size of the reference panel and its genetic relatedness to the study sample. We have recently proposed a general framework for the evaluation and investigation of gene-based rare variant tests of association. We have now extended this framework to incorporate rare variant genotype uncertainty and error models to precisely relate factors affecting genotype uncertainty to the power and type I error rate of rare variant tests. This work provides a realistic framework for assessing power and type I error and suggests genetic disease model, study design (e.g., imputation? sequencing depth?) and test statistic combinations that are particularly impacted by or resistant to genotype uncertainty.

1750F

Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics. Y.J. Hu1, S.T. Berndt2, S. Gustafsson3, A. Ganna4, J. Hirschhorn5,6, K.E. North5,7, E. Ingelsson8,9, D.Y. Lin10. Genetic Investigation of ANthropometric Traits (GIANT) Consortium, 1) Biostatistics and Bioinformatics, Emory University, Atlanta, GA. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, 30322, USA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA; 7) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 8) Divisions of Genetics and Endocrinology and Center for Basic and Translational Obesity Research, Children's Hospital, Boston, Massachusetts 02115, USA; 6) Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA; 7) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 8) Divisions of Genetics and Endocrinology and Center for Basic and Translational Obesity Research, Children's Hospital, Boston, Massachusetts 02115, USA; 6) Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA; 7) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 8) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, 30322, USA; 3) Winona State University, Winona, MN; 4) Brown University, Providence, RI, 5) Miami University of Ohio, Oxford, OH; 6) Dordt College, Sioux Center, IA.

Meta-analysis of genome-wide association studies (GWAS) has led to the discoveries of many common variants associated with complex human diseases. There is a growing recognition that identifying causal rare variants also requires large-scale meta-analysis. The fact that association tests with rare variants are performed at the gene level rather than at the variant level poses unprecedented challenges in the meta-analysis. First, different studies may adopt different gene-level tests, so the results are not comparable. Second, gene-level tests require multivariate statistics (i.e., components of the test statistic and their covariance matrix), which are difficult to obtain. To overcome these challenges, we propose to perform gene-level tests for rare variants by combining the results of single-variant analysis (i.e., p-values of association tests and effect estimates) from participating studies. This simple strategy is possible because of an insight that multivariate statistics can be recovered from single-variant statistics, together with the possibility of combining the results of single-variant analysis. We show both theoretically and numerically that the proposed meta-analysis approach provides accurate control of the type I error and is as powerful as combining the results of single-variant analysis. We compute gene-level test statistics and perform meta-analysis of any disease phenotype and any study design and produces all commonly used gene-level tests. An application to the GWAS summary results of the Genetic Investigation of ANthropometric Traits (GIANT) consortium reveals rare and low-frequency variants associated with human height. The relevant software is freely available.

1751W

Reinforcement policy for select-based association analysis. H. Huang1, P.O. Wu2, F. Xu3, Y. Hu1. 1) Abt Associates, North Chicago, IL; 2) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA. School of Public Health, University of Washington, Seattle, WA.

The advent of next-generation sequencing technologies affords the ability to sequence thousands of subjects cost-effectively, and is revolutionizing the landscape of genetic research. With the evolving genotyping/sequencing technologies, it is uncertain to expect that we will soon obtain a pair of high-quality fully phased genome sequences from each subject in the near future. However, in this context, we propose a novel framework called, recursive organizer (ROR), which recursively groups sequence variants based upon sequence similarities and their empirical disease associations, into fewer and potentially more interpretable super sequence variants (SSV). As an illustration, we applied ROR to assess an association between HLA-DRB1 and type 1 diabetes (T1D), discovering SSVs of HLA-DRB1 that are associated with T1D, a four-fold reduction of sequence complexity. Using HLA-DRB1 data from Type 1 Diabetes Genetics Consortium (T1DGC) as cases and data from Fred Hutchinson Cancer Research Center as controls, we were able to validate associations of these SSVs with T1D. Further, SSVs consist of nine nucleotides, and each associates with its corresponding amino acids. Detailed examination of these selected amino acids reveals their potential functional roles in protein structures and possible implications to the mechanism of T1D.
Valid permutation tests for genetic case-control studies with missing genotypes. D.D. Kinnamon, E.R. Martin. John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

Permutation tests are indispensable tools in the analysis of genetic case-control studies. However, many permutation tests proposed in the genetics literature are justified only by heuristic arguments under the assumption of complete genotype data. We therefore developed a rigorous theoretical framework for constructing valid permutation tests for genetic case-control studies with unrelated subjects and missing genotypes arising from a variety of missing data processes. We began by specifying a nonparametric probability model for the observed data in such a study. Using group-theoretic arguments, we then established two conditions that together guarantee an exact level-\( \alpha \) Monte Carlo permutation test for data generated under this nonparametric probability model. We showed that one of these conditions is not satisfied for the most frequently used Monte Carlo permutation test and that this test is guaranteed to be level \( \alpha \) only for missing data processes with certain characteristics. We therefore proposed an alternative Monte Carlo permutation test that we showed is exact level \( \alpha \) as long as all covariates influencing the missing data process are identified and recorded. We supplemented our theoretical development with Monte Carlo simulations for a variety of test statistics and missing data processes. We concluded that, while permutation tests can be extremely useful in genetic case-control studies, they must be constructed with careful consideration of the process generating missing genotypes to avoid inferential errors.

**1753F**

Heritability of Gene Expression Variation. Z. Liu\(^1\), J. Huang\(^1\), W. Cookson\(^2\), M. Moffatt\(^2\), L. Liang\(^3\). \(^1\) Department of Epidemiology, Harvard School of Public Health, Boston, MA; \(^2\) National Heart and Lung Institute, Imperial College, London SW3 6YJ, UK.

We proposed a statistical approach to estimate the heritability of global gene expression variability, named variance heritability (abbreviated in this paper as \( \text{vh2} \)), in Epstein-Barr virus-transformed lymphoblastoid cell lines in two large datasets from nuclear families of British descent recruited through a proband with childhood eczema (MRCE, \( n=550 \), based on Illumina platform) and asthma (MRCA, \( n=405 \), based on Affymetrix platform). This \( \text{vh2} \) estimates additional genetic effects on expression to those captured by narrow sense (additive) heritability, including dominant effect, GxG and GxE interactions. The median heritability of global gene expression variability across all probes in MRCE and MRCA were 0.04 and 0.04, the third quartile (Q3) were 0.12 and 0.14, respectively. Thirty percent of probes in MRCE and 34% of probes in MRCA have estimated \( \text{vh2} \) larger than 0.1. Compared with estimates from randomly permuted data (median=0.004, Q3=0.08, 20% probes with \( \text{vh2}>0.1 \) in MRCE; median=0.002, Q3=0.09, 22% probes with \( \text{vh2}>0.1 \) in MRCA), our estimated results implied the existence of genetic variants regulating the variance of gene expression. For probes targeting the same transcript, the heritability estimates were consistent across the two independent datasets (MRCE and MRCA) and were robust to expression array platforms. We further mapped the heritability of gene expression variance to individual genetic loci using Illumina 300K SNPs and 39 millions imputed 1000Genomes SNPs and INDELS. Our results suggest new genetic regulatory variants that might be missed by standard eQTL mapping studies that focused on the mean level of gene expression rather than its variability.

**1754W**

On the simultaneous testing for large genomic regions: A clustering approach for rare variants. S. Lutz\(^4\), C. Lange\(^5\). \(^1\) Biostatistics, University of Colorado, Aurora, CO; \(^2\) Biostatistics, Harvard School of Public Health, Boston, MA.

In case-control studies, we propose a general analysis frame-work in which hundreds of genetic loci with allele frequency less than 1% can be tested simultaneously for association with case-control status. The approach is built on spatial-clustering methodology, assuming that rare variants that are associated with the target phenotype cluster in certain genomic regions. In contrast to standard methodology for rare-variant analysis, which has focused on collapsing methods, the proposed approach does not depend on each rare variant being associated with the case-control status in the same direction and with similar effect sizes. Using simulation studies, the properties of the approach are evaluated. In an application to a genome-wide association study (GWAS) for chronic obstructive pulmonary disease (COPD) in the COPDGene study, we illustrate the practical relevance of the proposed method.

**1755T**


Technological advances permit us to collect data on thousands of single nucleotide polymorphisms (SNPs) for each individual in a genetic association study. However, current paradigm of analyzing such dataset focuses on looking at association with the phenotype only at a single SNP level. Due to limited success of genome wide association study using this method, some multilocus association methods have been proposed in recent literature. Even then an important issue is often ignored while collecting such data. The phenotype data, whether qualitative or quantitative, is usually collected only once. However, due to medication and other factors like age, food habits etc the quantitative phenotype value may vary over time. Thus a longitudinal phenotype data i.e. repeated measurements on the quantitative phenotype of an individual at different time points might be more informative to get a more vivid picture of the genetic architecture. In this work we have developed a multilocus genetic association testing method using longitudinal quantitative phenotype data to see any genetic association between any genomic region of interest or genes (rather than a single SNP) with the phenotype. Our method explores the association between kernel based genotype similarity and phenotype similarity over different time points and combine information of such association over multiple markers. This method is flexible enough to adjust the effect of other covariates, if any and also can takes care of the missing value for the covariates at some time points. We have also derived the asymptotic distribution of the test statistic that will help to calculate the \( p \)-value faster than a permutation technique.

**1756F**

Adaptive approaches for combining multiple rare variant association tests provide improved power across a wider range of genetic architecture. M. O’Connell\(^1\), A. Benitez\(^2\), J. Arbet\(^3\), K. Grindle\(^4\), C. Fu\(^5\), B. Greco\(^6\), A. Hainlin\(^7\), N. Tintle\(^8\). \(^1\) Miami University of Ohio, Oxford, OH; \(^2\) Brown University, Providence, RI; \(^3\) Winona State University, Winona, MN; \(^4\) St. Olaf College, Northfield, MN; \(^5\) Massachusetts Institute of Technology, Cambridge, MA; \(^6\) Dordt College, Sioux Center, IA; \(^7\) Vanderbilt University, Nashville, TN; \(^8\) University of Michigan, Ann Arbor, MI.

Over the past five years, numerous gene-based rare variant tests of association have been proposed, each of which attempt to combine variants within a gene or region of interest into a single association statistic, with a goal of providing more power than a strategy which analyzes each variant separately. Simulation results have shown that many of these individual tests provide good power for particular genetic architectures, but not others. We have developed a general strategy for combining any two or more gene-based rare variant tests using an adaptive approach, which yields a single \( p \)-value representing the cumulative evidence for association across the set of gene-based tests. For example this strategy can take any threshold based test and turn it into a variable-threshold test, combine similar tests (similar statistic with alternative weighting strategies), or combine substantially different tests (e.g., burden tests and variance components tests). Using simulation we provide guidance on the tradeoff between power gains and test robustness versus the number of tests being combined, a result which is based on the correlation structure of the tests are under the null hypothesis of no association. Finally, we demonstrate how recent results from our group which suggested a substantially different gene-based test which is robust to high proportions of non-causal variants, combined with other popular tests (burden and variance component tests), can provide improved power across a wider range of genetic architecture.
Posters: Statistical Genetics and Genetic Epidemiology

1757W
Genome-wide association study of allergic rhinitis, C. Schaefer1, J. Liu1,2, L. Shen1, T. Hoffmann1,2, M. Kvalø1,2, Y. Banda1,2, D. Ranatunga4, N. Risch1,2, J. Wilte1,2, E. Jorgenson1,2, 1) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 3) Kaiser Permanente Northern California Division of Research, Oakland, CA;

Allergic diseases are complex and the risk of developing allergies is thought to have a strong genetic basis. To examine the role of genetic factors on allergic rhinitis, we conducted a genome-wide association study in 61,957 non-Hispanic White subjects from the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, which includes genotypes on over 675,000 SNPs. We identified several variants that were associated with allergic rhinitis at a genome-wide significant level. The most significant SNP associations identified are located on chromosome 2 within the gene IL1RL1 (rs2160203, p=3.8×10⁻10). IL1RL1 encodes a member of the interleukin 1 receptor family and has previously been reported to be associated with asthma. In addition, three SNPs (rs3806933, rs34624588, and rs14386737) that lay within a region of chromosome 5 containing the gene TSLP and WDR36 reached genome-wide significance (5×10⁻8). TSLP promotes Th helper 2 cell responses and is associated with immune response in various inflammatory diseases, including asthma and allergic inflammation. WDR36 is involved in T-cell activation and has been shown to be associated with bronchial asthma and eosinophilic esophagitis. In summary, we detected genome-wide significant SNP associations for allergic rhinitis, in loci that have been previously reported to be associated with asthma and allergic disease.

1758T

Association studies that genotype affected offspring and their parents (triads) offer robustness to genetic population structure while enabling analyses to probe maternal effects, parent-of-origin effects, fetal-maternal interaction, and gene-by-environment interaction. We study case-parents designs that use pooled DNA specimens to both reduce genotyping costs and make good use of limited available specimens. We assume that the assay for each diallelic autosomal locus counts the number of variant alleles among the 2h copies present in a pool of h individuals. We reduce the genotypes required per pool from 3h to three by randomly partitioning the available triads into pooling sets of h triads each and then creating three pools from every pooling set, one pool each for mothers, fathers, and offspring. Data analysis proceeds via log-linear modeling using the expectation maximization algorithm, where the pseudo-complete data are disaggregated allele counts. The analysis provides relative risk estimates and assessment of maternal and parent-of-origin effects. It also accommodates genotyping errors and missing genotypes. We compare the power of our proposed analysis for testing offspring and maternal genetic effects to that based on a difference approach considered by Lee and that of the gold-standard based on individual genotypes, under a range of allele frequencies, missing father proportions, genotyping error rates and modes of inheritance. Power calculations show that the pooling strategies incur only modest reductions in power if genotyping errors are low, while reducing genotyping costs and conserving limited specimens. We illustrate our procedure with data from a study of oral clefts.

1759F
Meta-Analysis of Genome-Wide Association Studies in Myopia in Nine Populations, C. Simpson1, R. Wojciechowski2, V. Verhoeven3,4, P. Hysi5, M. Schadt6, V. Li7, M. Humair8,9, X. Fu10, F. Vorbs10,2, F. Murgia10, K. Oexle1,11,2, A. Paterson1,2, V. Vitart3, C. Hammond4, P.N. Baird4, M. Pirastu10, J. Rotter2,3, C.C.W. Klaver2,4, T. Meiling1,12, D. Stambolian1, J.E. Bailey-Wilson1,3, 1) Inherited Disease Research Branch, NHGRI, NIH, Baltimore, MD; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA; 3) Department of Ophthalmology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands; 4) Department of Epidemiology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands; 5) Department of Twin Research and Genetic Epidemiology, King’s College London, St. Thomas’ Hospital, London, UK; 6) Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, Melbourne, Australia; 7) Medical Genetics Institute, Cedars-Sinai Medical Center, CA, USA; 8) Institute of Medical Science, University of Toronto, Toronto, Canada; 9) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Canada; 10) Institute of Population Genetics, National Research Council, Sassari, Italy; 11) Helmholz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology I, Neuherberg, Germany; 12) Institute of Human Genetics, Technical University Munich, Munich, Germany; 13) Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; 14) Department of Ophthalmology, University of Pennsylvania, Philadelphia, USA.

Myopia is a common refractive error which affects at least a third of most populations. Both genetic and environmental factors influence myopic development. It has a significant impact on the lives of affected individuals and carries high economic costs associated with treatment, loss of productivity and co-morbidity from vision impairment. Recent genome-wide association studies (GWAS) have identified a number of loci associated with myopia and refractive error. Here we report results of a large meta-analysis of myopia in nine cohorts, for a total of 17,787 individuals of European ancestry and replication in a further 8 cohorts for a total of 7953 individuals. Genotypes in each population were imputed to HapMap2 and analyzed separately by each genotype. Cases were defined as a spherical equivalent of ≥1 diopters (D) or worse and controls were defined as > 0D. Individuals between 0 and -1D were coded as unknown. Analyses were performed including age, sex and years of education, plus the first 3 principal components to adjust for population structure. Metas-analysis for a total of 20,431 SNPs. The replication threshold was set by calculating the effective degrees of freedom using the sample size formula recommended by the authors of METAL. Genomic control was used to adjust for any residual structural differences between populations. 3 SNPs were identified as genome-wide significant with P<5e-6, rs10113215 on chromosome 8q12.1 and rs1370156 and rs2028099 both in a previously reported locus on chromosome 15q14. For replication, SNPs with P<1e-5 were identified and all SNPs within 500kb each side of that SNP selected, for a total of 20,431 SNPs. The replication threshold was set by calculating the effective degrees of freedom using the Ramos method. SNPs will be considered to replicate where the p value < 0.0026. The replication analyses are ongoing and will be presented.

1760W
Identifying disease susceptibility variants using pleiotropy and independent genome-wide association studies. A. Skol. Genetic Medicine, Dept of Medicine, Univ Chicago, Chicago, IL.

Pleiotropy is the phenomenon whereby a single gene influences multiple traits. Given that observing more than one disease cosegregating within a family is not uncommon, and that several studies have mapped the same gene to multiple traits, it would appear that pleiotropy is not an unusual phenomenon in disease genetics. Here, I explore the ability to use pleiotropy to identify variants or genes that are common between diseases using independent GWAS data, meaning that the traits of interest are measured in independent groups of individuals. The rationale for looking for pleiotropic loci is two-fold: first, most studies will be insufficiently powered to identify most of the trait or disease influencing variation, however if two traits share a genetic underpinning then we should find contributing variants in the intersection of the studies’ marginally significant results, and second, pleiotropic loci will provide novel insights into the common genetic etiology of seemingly disparate disease. I will compare three methods for detection: meta-analysis, intersection of the tails of the GWAS test statistics, and false discovery rate of p-values from one GWAS within the tail of the other. I will present results demonstrating the power of these approaches, how to determine significance thresholds that maximize power, what effect sizes are detectable, and how to correctly accommodate linkage disequilibrium.
1761T An extrapolation approach for estimating genome-wide significance for whole genome sequencing studies and region-based tests. C. Xu1,2 on behalf of the UK10K Statistics group. 1) Lady Davis Institute, JGH, Montreal, Canada; 2) Dept. of Biostatistics, McGill University.

Rare genetic variation derived from whole genome sequencing studies is often analyzed by defining a series of windows or genetic regions, and then evaluating the association between phenotype and all the genetic variability in each region. However, the choice of region boundaries, test statistics, and model parameters for such tests may be somewhat arbitrary, and repeated analyses may be performed making different assumptions. Therefore, establishing an appropriate correction for multiple testing is a challenging problem. In recent work, we have used pilot data from the UK10K consortium whole genome sequencing study to propose an empirical approach to establishing significance thresholds. This is combined with extrapolation from small genomic regions to the whole genome to make the empirical evaluations computationally feasible. Here we evaluate the dependence of the resulting estimated significance thresholds on sample size and minor allele frequency. In our investigations, we show that significance thresholds, for a combined analysis including univariate tests for common SNPs and region-based tests for rare genetic variation, are driven by the common SNP variation, and that we saw little dependence of the results on sample size. We recommend a genome-wide significance threshold of $10^{-6}$ for such a combined analytic strategy.


Recent advances in exome sequencing and the development of exome genotyping arrays are enabling explorations of the contribution of rare variants of clear functional consequence to complex traits. Decreasing cost of sequencing and the low cost of exome chip genotyping allow increasingly large cohorts to be studied. For these rare variants, single variant association tests can lack power and a variety of association tests that group rare variants by gene or functional unit have been proposed as alternatives. Here, we describe family-based association tests for rare variants that allow analysis of a variety of quantitative traits, with or without covariates, and show how these tests can be applied in meta-analysis settings. The key idea beyond our method is that various gene-level test statistics can be reconstructed from single variant score statistics and that, when the linkage disequilibrium relationships between variants are known, the distribution of these gene-level statistics can be derived and used to evaluate significance. This idea can be used to derive a variety of gene-level association tests in families and also to enable meta-analysis. We have implemented family-based burden test, SKAT and variable frequency threshold tests. The methods have been implemented in freely available C++ code. To investigate power, we simulated families with various structures and sizes and genes comprising between 1kb and 25kb of sequence, each explaining 1-5% trait variance. In each gene, 10-30% rare variants were chosen to be causal. Our simulations show that, using alpha=2.5x10^{-6}, when all variants influence the trait in the same direction, the variable threshold model has the largest power. For a sample of three-generation-family of size 10 with 1000 founders or 2500 individuals, VT is predicted to obtain ~80% power when 80% variants are causal for a gene of 1kb explaining 2% of trait variance. When variants influencing traits in both directions co-exist in the same gene, half of them are trait-decreasing, our implementation of the SKAT test performs the best. For a sample of three-generation-family of size 16 with detected founders or 8000 individuals, SKAT is predicted to obtain ~85% power when 80% variants are causal and 50% of them are trait decreasing for a gene of 1kb explaining 4% of trait variance. Type-I error was examined using simulations under the null hypothesis and they were well controlled.

1763W Testing for differences between multiple groups in high-throughput sequencing data using Bayesian multi-scale Poisson models. H. Shim1, E. Pantaleo1, M. Stephens1,2, 1) Dept Human Genetics, Univ Chicago, Chicago, IL; 2) Dept Statistics, Univ Chicago, Chicago, IL.

High-throughput sequencing technologies are now routinely applied at a genome-wide scale to collect a variety of phenotypic data. Testing for differences in these data between multiple groups is frequently encountered in genetics applications (e.g., eQTL mapping using RNA-seq and detecting differences in transcription factor binding across tissues using DNase-seq or ChiP-seq). Most approaches perform a test for differences using summary statistics such as total number of reads mapped to a gene or window although the original data consists of the counts of reads mapped to each base along the genome. These typical approaches have limited power to identify diverse patterns of signals and their results are often sensitive to window size. For example, these approaches could miss associations when different bases within a window show effects in opposite directions and partially cancel each other out, leading to a small overall effect on total number of reads over the window. Here we present statistical methods for testing for differences in the original data, which leads to increased power by using full information from the data. Specifically, our methods consider the data as an inhomogeneous Poisson process and test for differences in underlying intensities using Bayesian multi-scale Poisson models. In addition to testing for differences, our approach aims to provide a better interpretation of the analysis such as which parts and features of the data are driving the observed signals. We illustrate the proposed methods on DNase-seq data from 70 HapMap Yoruba LCLs. Moreover, we modify the proposed method to detect differences in gene expression between multiple tissues measured by paired-end RNA-seq data. Our results demonstrate that the proposed approaches can considerably increase power to detect associations compared with conventional window-based approaches.

1764T Detecting association of rare variants by testing an optimally weighted combination of variants for censored survival outcomes. X. Wang, X. Zhao. Joseph J. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI. 1240 N 10th Street Milwaukee, WI 53205.

There is increasing evidence showing that rare genetic variants are important for identifying patients with aggressive diseases who need prioritized treatments. Identifying these patients is critical in improving survival outcomes. In addition, knowledge of the rare genetic variants can help us gain further understanding the biological processes underlying aggressive diseases progression. Next-generation sequencing technology allows sequencing the whole genome of large groups of individuals with survival outcomes, and thus it makes directly testing diagnostic rare variants possible. However, recently developed statistical methods for detecting association of rare variants are not applicable to survival data. Survival data are often subject to censoring and use the Cox proportional hazards model. The score test in a Cox proportional hazards model is seriously anti-conservative for rare variants. In this paper, we use the signed square root of the log partial likelihood ratio test statistic to replace the score test and propose a Variable Weight Test for testing the effect of an optimally weighted combination of variants on a survival outcome. Our new test is based on optimal weighted rare variants and is a weighted score test. We perform extensive simulation studies show that the type I error rates of the new test are under control. In order to evaluate the performance of the new test, we are conducting extensive simulation studies for power comparisons. Furthermore, we will illustrate the proposed test with a real data application.

1765F Detecting association for low-frequency variants by the standardized linkage disequilibrium in case-control genome-wide screens. C. Xing, C-Y. Lin, H-C. Ku. McDermott Ctr, Univ Texas SW Med Ctr, Dallas, TX.

In genetic association studies a conventional test statistic is proportional to the correlation coefficient between the trait and the variant, with the result that it lacks power to detect association for low-frequency variants. Considering the link between the conventional association test statistics and the linkage disequilibrium patterns, we propose a test statistic analogous to the standardized linkage disequilibrium D-primer to enhance the power of detecting association for low-frequency variants with moderate to large effect sizes. We examined its validity and showed it is more powerful than the conventional test statistic when rare variants are not in linkage equilibrium. We evaluated the power of detecting association for low-frequency variants in a genome-wide screen setting by both simulation and real data analysis.
1766W
An optimal quasi-likelihood-based burden test for rare-variant association. X. Wu1, H. Zhu1, D. Liu1. 1) Department of Statistics, Virginia Tech, Blacksburg, VA; 2) Department of Bioinformatics, University of Michigan, Ann Arbor, MI.

With the advent of next-generation sequencing (NGS) technologies, rare-variant association testing is receiving increasing amount of attention in genome-wide association studies. As a result, statistical tests for detecting rare variants (single nucleotide polymorphisms, SNVs) have been developed. However, a critical limitation of these tests is that they are powered to detect only one variant per locus. As a result, many variants may not be appropriate in the presence of related individuals. Furthermore, the weighting scheme in those burden tests is often empirical and lacks theoretical justification. In this paper, we propose a novel MQLS burden test (OMBET), a flexible and efficient method for testing multiple-variant (rare or common) association with disease-related traits (quantitative or dichotomous). OMBET is a generalization of the traditional MQLS/MASTOR test in order to accommodate rare variants under linkage disequilibrium (LD). Compared with other burden tests, OMBET has several advantages: (1) it takes into account dependence among individuals and among variants, (2) it can incorporate individuals with missing data, and (3) it provides optimal weights to guarantee maximized power for association testing. Simulation studies demonstrate that OMBET achieves correct type-I error and improved power over other competing methods. We apply the proposed method to a sequencing-based genetic study using data from the SardiNIA Medical Sequencing Discovery Project.

1767T
An empirical validation of random effects and Bayesian meta-analysis models. R. Ahn, C. Garner. Department of Epidemiology, Sprague Hall, Room 318, University of California, Irvine, Irvine, CA 92697-3905.

Genome-wide association study (GWAS) meta-analysis is a statistical method that is now routinely used to test hypotheses drawn from individual data or summary statistics from multiple studies to increase the power to detect the small genetic effect sizes of common alleles and to decrease the number of false-positive associations. As new samples do not have to be genotyped, the cost to perform a GWAS meta-analysis with sample sizes in the tens of thousands or even in the hundreds of thousands is substantially lowered. As a result, hundreds of GWAS meta-analyses have been performed over the last few years and have substantially increased the number of risk loci discovered and replicated for several different phenotypes. Amongst the many approaches to GWAS meta-analysis, the most popular has been the fixed effects meta-analysis because it is the most powerful approach. However, fixed effects meta-analysis ignores the potential heterogeneity that may exist between studies by assuming that the effect of a risk allele is the same across all studies. We have reanalyzed a two-stage GWAS meta-analysis using 292K single nucleotide polymorphisms (SNPs) directly genotyped in 4,533 celiac disease cases and 10,750 controls from 4 different countries in the discovery stage and 131 SNPs were directly genotyped in 9,451 celiac disease cases and 16,434 controls from 11 countries. In the original study, a fixed effects meta-analysis approach was used. The purpose of our analysis was to determine if a novel random effects or Bayesian meta-analysis model that accounted for between-study differences could yield results that differ from the results that were originally published. Here we present evidence that a SNP at a loci in chromosome 1 (RUNX3) that was previously reported to show genome-wide significance (P < 5 x 10^-8) was not genome-wide significant in either the original or the reanalysis. However, when the between-study heterogeneity was accounted for by the random effects or Bayesian meta-analysis approaches, this study highlights the need to carefully investigate between-study heterogeneity and the implementation of GWAS meta-analysis models that control for heterogeneity.

1768F
Methods and tools for fast efficient mixed-models based whole-genome association analysis for large cohorts and multiple phenotypes. Y. Aulchenko. Lab of recombination and segregation analysis, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russian Federation. Lab of recombination and segregation analysis, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russian Federation.

Whole-genome association (WGA) analysis is a tool of choice for identification of loci underlying complex human traits. In WGA scans, association of millions of genetic polymorphisms measured with SNP arrays and/or genome sequencing (e.g. deep resequencing) is performed with the trait among thousands of individuals. The mixed models methodology allows powerful WGA analysis in case when the trait under analysis is highly polygenic and/or when genetic (sub)structure is present in the sample. However, the analysis under mixed models is computationally challenging, especially for large cohorts and/or multi-trait “omics” data sets. Here, I will review recent advances in mixed-models based WGA analyses, and describe a number of methods, algorithms, and software tools we have developed to facilitate the analysis of cohorts including tens of thousands of participants and analysis of “omics” data sets potentially including hundreds of thousands of phenotypes.

1769W
PODKAT: a non-burden test for associating complex traits with rare and private variants. U. Bodenhofer, S. Hochreiter. Institute of Bioinformatics, Johannes Kepler University, Austria.

High-throughput sequencing technologies have facilitated the identification of large numbers of single-nucleotide variations (SNVs), many of which have already been proven to be associated with diseases or other complex traits. Since association signals with SNVs are known to be underpowered, different collapsing strategies have been proposed to consider multiple SNVs occurring in a region simultaneously. Such strategies can be classified into burden tests and non-burden tests, an important representative of which is the acclaimed Sequence Kernel Association Test (SKAT) by Wu et al. Several large sequencing studies, such as the 1000 Genomes Project, the UK10k Project, or the NHLBI-Exome Sequencing Project, have consistently reported a large proportion of private SNVs, that is, variants that are unique to a family or even a single individual. Non-burden tests like SKAT are typically utilizing correlations between SNVs to increase statistical power - a strategy that is not applicable to private SNVs, since singular events are generally uncorrelated. Burden tests are potentially able to deal with private SNVs, but only if the number of private SNVs occurring in a region is correlated with the trait under consideration. Moreover, burden tests have a disadvantage if deleterious and protective SNVs occur together in the same region. We propose the Position-Dependent Kernel Association Test (PODKAT). By means of a position-dependent kernel approach, PODKAT can potentially detect associations of rare and private SNVs with the trait under consideration even if the burden scores are not correlated with the trait. PODKAT assumes that, the closer two SNVs are on the genome, the more likely they have similar effects on the trait under consideration as long as neither is deleterious, neutral, and protective variants are grouped sufficiently well along the genome. PODKAT can be applied to testing focused regions as well as to whole-exome and whole-genome association testing. We evaluated PODKAT on simulated data with single quantitatively trait (contingent and dichotomous) and real sequencing data with simulated and real traits to illustrate its potential for association testing involving rare and private variants.

1770T
Model-based pathway enrichment analysis and prioritization of genetic variants in enriched pathways yields novel putative susceptibility loci for rheumatoid arthritis and type 1 diabetes. P. Carbonetto1, M. Stephens2. 1) Dept. of Human Genetics University of Chicago Chicago, IL; 2) Depts. of Statistics and Human Genetics University of Chicago Chicago, IL.

Many common diseases appear to be highly polygenic, modulated by a large number of genetic factors each with a small effect on disease risk. As a result, standard single-marker analyses of genome-wide association studies are unable to identify most variants conferring risk to disease; many individual SNP contributions to disease risk are so subtle that they cannot be distinguished from distortions that occur by chance. To address this problem, we have developed statistical methods that aggregate information over groups of related genes, such as genes in common pathways, to identify gene sets that are enriched for variants associated with disease. However, these methods often fail to answer a critical question: which genes and variants in the enriched pathways are associated with disease risk? We have developed a model-based approach that not only systematically interrogates pathways for enrichment of disease associations, but also uses this information to prioritize variants assigned to these pathways. As an attempt to promote these variants above background noise, a key feature of this approach, which distinguishes it from other enrichment methods, is that it is able to uncover additional loci contributing to disease risk. We describe our results from applying this approach to genome-wide studies of Crohn’s disease (CD), rheumatoid arthritis (RA) and type 1 diabetes (T1D). These analyses yield many additional putative associations compared to standard single-marker analyses. Because of the predominant role of the MHC in RA and T1D, we developed methods to assess enriched pathways beyond enrichment of disease associations in the MHC, and found that this was important for reliably identifying pathways, and non-HLA associations. For CD and RA, most of the additional disease associations (7 of 8 non-HLA associations) are corroborated by other studies and large-scale meta-analyses, validating the usefulness of the methodology. The one non-HLA association identified appears to be a promising candidate. For T1D, the results very strongly support the connection between IL-2 signaling and development of T1D, and prioritization of IL-2 signaling genes yields strong evidence for 7 additional non-HLA disease risk loci. As disease risk has been validated by other genome-wide studies of T1D (IL2, IL2RA, CLEC16A and C10orf56), and 3 constitute novel T1D loci (regions containing RAF1, MAPK14 and FYN).
Rapid linear mixed model methods for large-scale genome-wide association studies. W-M. Chen1, A. Manichaikul1, S.S. Rich1, M. Cushman2, M.M. Sale1. 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Department of Medicine, University of Vermont, Colchester, VT.

Population stratification is known to confound association results in genome-wide association studies (GWASs). Recently, Linear Mixed Models (LMMs) have received attention as a flexible way to adjust for population structure using variance components. Multiple LMM algorithms were recently published that included exact maximum likelihood methods with computational complexity quadratic to sample size. Although existing LMM methods have proven to be computationally feasible for GWAS scans, several limitations exist. First, current methods have inflated type I errors in the presence of missing genotypes and mis specification of polygenic estimators; second, current efficient algorithms for LMM require analyzing one trait, limiting the number of traits that can be analyzed in a reasonable amount of time. Here, we present two novel fast LMM methods with computational complexity linear to sample size: (1) an approximation fast LMM method, and (2) an exact fast LMM method. As score tests, the proposed LMM test statistics are developed to be robust to missing genotypes, missing phenotypes, phenotypic distributional assumptions, and misspecification of polygenic parameters. The computational improvement of our implementations over existing fast LMM methods includes efficient computation of the relatedness matrix using the KING algorithm, rapid regression-based estimation for polygenic analysis in the approximation LMM method, and rapid genome scans in a matter of seconds for both fast LMM methods. In a test dataset consisting of 2,400 samples each typed at ~800,000 SNPs and 1,000 quantitative traits, it takes 40 minutes to perform 1,000 GWAS scans (i.e., 2 seconds per GWAS scan) using the approximation fast LMM method and 9 hours using the exact fast LMM method on a single CPU. The genome scan results using the proposed fast LMM methods are highly correlated (e.g., with correlation 0.994) with the results using the exact maximum likelihood methods, with the same order of magnitude of significance at the top GWAS hits. Efficient implementations in our software tool KING makes it feasible to conduct small GWAS scans for 10,000s of traits with small amount of computing resources.

An effective association testing procedure incorporating admixture mapping information. G. Gao, W. Chen. Department of Biostatistics, Virginia Commonwealth University, Richmond, VA.

Admixed populations are formed by recent admixture of two or more ancestral populations. For instance, African Americans often have recent genetic ancestry from both West Africans and European Americans. As a result of admixture, the variation in genome-wide ancestry (a type of stratification) in admixed populations can be a confounding factor in association tests, causing false positive and false negative findings. This makes genome mapping more complicated in admixed than in non-mixed populations (such as Europeans). For gene mapping in admixed populations, admixture mapping tests use admixture linkage disequilibrium (LD) and can only identify a causal variant in a large chromosomal region (several Mbs). Admixture LD occurs during admixture over the past several hundred years when large chromosomal segments are inherited from a particular ancestral population. To identify a causal variant in a small region, association tests that correct for local ancestry have been developed. These tests can use the background population (LD) existing in the ancestral populations and therefore can often identify a causal variant in a region with less than a few hundred Kbs. However, these tests can have relatively low power. Recently, to acquire increased power, several joint association tests that combine information from admixture mapping tests and association tests that correct for local ancestry have been proposed. However, these joint tests are more appropriate to identify a causal variant in a large chromosomal region (several Mbs). Our simulation studies showed that these joint tests could not control family-wise error rates (FWERs) in genome-wide association studies if the null hypothesis was true. In admixed populations, the null hypothesis is not true in any case if a SNP is causal. In order to understand the genetic architecture of complex diseases it is therefore crucial to unravel the genetic dependencies between correlated disease traits (pleiotropic effects).

Mixed model association statistic with correction for case-control ascertainment provides large increase in power. T. Hayeck1, N. Zaitlen2, B. Viljanen3, S. Pollack1, J. Yang2, G. Chen3, M. Goddard4, P. Visscher1, N. Patterson2, A. Price1. 1) Harvard School of Public Health, Boston, MA; 2) University of California, San Francisco; 3) University of Queensland, Brisbane, Australia; 4) University of Melbourne, Melbourne, Australia; 5) Broad Institute, Cambridge, MA.

We introduce a mixed model association method for ascertained case-control studies that increases test statistics at causal markers by up to 23% vs. existing mixed model methods, with well-controlled false-positive rate. It is widely known that appropriate modeling of case-control ascertainment can produce large increases in power for case-control studies with fixed-effect covariates (reviewed in Mefford & Witte 2012 PLoS Genet), but such increases in power have not yet been obtained for models that include random effects. Here, we improve upon existing mixed model methods (e.g., Kang et al. 2010 Nat Genet, Zhou & Stephens 2012 Nat Genet) using a score statistic computed from posterior mean liabilities (PML) under the liability threshold model. The PML of each individual is conditional not only on that individual’s case-control status, but also on every other individual’s case-control status and on the genetic relationship matrix obtained from the data. For example, disease cases with higher genetic relationships to other disease cases will be assigned a larger PML than disease cases with lower genetic relationships to other disease cases. The PML are estimated using a multivariate Gibbs sampler, with the liability-scale phenotype covariance matrix based on the genetic relationship matrix and a heritability parameter estimated via Haseman-Elston regression on case-control phenotypes followed by transformation to liability scale (Lee et al. 2011 AJHG). The Gibbs sampler does not iterate over SNPs, and overall running time is comparable to existing mixed model methods. In simulations of unrelated individuals, our statistic outperformed existing mixed model methods in all scenarios tested, with the magnitude of the improvement depending on the severity of case-control at LMM methods. For example, we observed 23%, 8% and 2% improvements in test statistics at causal markers for ascertained case-control studies simulated at thresholds of 3, 2 and 1 (corresponding to disease prevalence of 0.13%, 2.5% and 16%), with well-calibrated test statistics and good control of false-positive rate. In addition, it makes it feasible to conduct large-scale GWAS scans for 10,000s of traits with small amount of computing resources.
1775W
Variable selection for GWAS with linear mixed models yields improved power and control of type I error. D. Heckerman¹, O. Weissbrod², N. Fusili¹, C. Kadie¹, R. Davidson¹, C. Lippert¹, J. Listgarten¹. 1) Microsoft Research, Los Angeles, CA; 2) Computer Science Department, Technion, Haifa; 3) Microsoft Research, Redmond, WA.

Linear mixed models (LMMs) are now routinely used to correct for genetic structure in genome-wide association studies (GWAS). At their core, LMMs rely on the estimation of a genetic similarity matrix, which encodes the pairwise similarity between every two individuals in a cohort. These similarities are estimated from single nucleotide polymorphisms (SNPs) or other genetic variants. Traditionally, all available variants are used to estimate the matrix. Here, we provide both theoretical and empirical evidence that such use is non-optimal, and that the careful selection of variants can lead to improved results. Theoretically, a linear mixed model is equivalent to a form of linear regression, where the SNPs that determine the genetic similarity matrix in the LMM view are covariates in the linear-regression view. Taking the latter view strongly suggests that the inclusion of variants relevant to the phenotype and the exclusion of variants irrelevant to the phenotype should lead to improvements in analysis. Empirical results confirm this argument, using a simple variant selection algorithm that searches over various sets of SNPs to identify those that maximize cross-validated prediction accuracy. To keep the search practical, we order SNPs for each fold by their univariate linear-regression $P$ values on the training data for that fold. We then use increasing numbers of SNPs by this ordering, measuring prediction accuracy on the out-of-sample test set. Next, we average the prediction accuracy over each fold. Finally, we identify the number of SNPs that optimized this average. Essential to this procedure is out-of-sample cross-validation. In this view, the number of parameters of the LMM represents the ratio of the variance explained by noise to the variance explained by the genetic similarity matrix. In-sample optimization of this parameter leads to lower quality GWAS performance. We apply this variant selection algorithm to several phenotype-genotype associations, including $2q11$ and WTCCC2 data, demonstrating significant, simultaneous improvements in control of type I error and power.

1776T
Association Studies of Imputed Genotypes Using Expectation-Maximization Likelihood-Ratio Test. K. Huang¹, Y. Li¹,². 1) Department of Biostatistics, The University of North Carolina, Chapel Hill, NC; 2) Department of Genetics, The University of North Carolina, Chapel Hill, NC.

Genotype imputation has become standard practice in modern genetic studies. As sequencing-based reference panel continues to grow, we have increasingly more well-imputed markers but at the same time also more imputation markers with relatively low quality. Here, we propose new methods that attempt to more elegantly incorporate uncertainty when analyzing imputed genotypes. We consider two scenarios: 1) when posterior probabilities of genotypes are estimated or 2) when only imputed dosages are available. When posterior probabilities are estimated, we developed an expectation-maximization (EM) likelihood-ratio test (LRT) for association studies. When only dosages are observed, instead of modeling dosages directly, we first sample the probabilities of all three possible genotypes and then apply the EM-LRT test on the sampled probabilities for EM-LRT. Extensive simulations have shown that type I error rates of the EM-LRT tests under both scenarios are protected. Regarding power, EM-LRT-Prob offers enhanced statistical power across the whole spectrum of imputed quality and EM-LRT-Dose has similar power performance as EM-LRT-Prob and better than standard methods that model dosages directly, especially for markers with relatively low imputation quality ($R^2 < 0.3$). Application to real datasets will also be shown.

1777F
Bayesian sparse models of high-dimensional correlated traits in related and unrelated individuals. V.itchkova¹, J. Connell⁴, A. Dah⁴, J. Marchini¹,²,⁴. 1) The EMLB-European Bioinformatics Institute, Hinxton, United Kingdom; 2) The Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 3) Department of Statistics, Oxford University, Oxford, United Kingdom; 4) The Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom.

Genetic association studies of complex traits have yielded a wealth of biologic discoveries. However, up to now studies have been mostly carried out under simplistic assumptions, for instance analyzing one trait and one SNP at the time, thus failing to capture the underlying complexity of these datasets. In order to move beyond simple GWAS the joint analysis of complex, highly-dimensional datasets represents an important extension of phenotype-genotype associations with great potential. The move to high-dimensional phenotypes raises many new statistical problems. For example, how to model the large diversity of possible different types of phenotypes, accounting for sporadically missing phenotype data, modelling sparse signals and hidden factors, how to model relationships and networks between phenotypes and how to assign statistical significance. We have developed a Bayesian model designed to detect association when correlations exist between high-dimensional phenotypes and between individuals due to relatedness. We impose a sparsity assumption through the use of a ‘spike and slab’ prior on the genetic effect sizes that allows us to learn the subset of traits that are associated with each SNP or genetic region. In this way our method is able to carry out model inference (association testing) and model selection at the same time. To allow for related individuals and/or protect against population structure our method generalizes single-trait linear mixed model analysis to high-dimensional traits. We use variational bayesian methods to efficiently fit our model, which facilitates approximate fully Bayesian inference. A key property of our method is that it can handle missing phenotype data, which we impute as part of the model fitting process. To illustrate the power of our method we have applied it to a study of high-dimensional glycomics phenotypes on 960 individuals from an isolated population where individuals show appreciable levels of relatedness. At the MGA5 association on 2q21 (rs1257220), which was originally discovered in a much larger sample, our method clearly detects an association (log10 Bayes Factor = 5.6), whereas approaches based on combining single-trait analyses uncover little signal. On simulated datasets of high-dimensional traits our approach has significantly more power than simpler approaches when the true signal is sparse, can accurately impute missing phenotypes and control for population structure and relatedness.

1778W
Effect size estimation in the stage of planning a replication study. C. Kuo¹, D. Zaykin². 1) Department of Community Medicine & Health Care, University of Connecticut, Farmington, CT; 2) Biostatistics Branch, National Institute of Environmental Health Sciences, NIH.

Large-scale association and sequencing studies generate millions of results. The top hits usually sorted by $P$-value are of interest but the smallest $P$-values can result from true signals or simply occur by chance. A common strategy to separate true effects from null effects is to carry forward a number of top hits to a replication study. In the replication study, the number of top hits can be chosen to control family wise error rate or false discovery rate. The sample size that depends on the effect sizes such as odds ratios or hits can be chosen to control family wise error rate or false discovery rate. For the winner’s curse mainly fall into ‘conditional likelihood approaches’ or ‘Bayesian approaches’. It has been criticized that the conditional likelihood approaches don’t use the population information provided by samples. The Bayesian approaches are immune to the winner’s curse while they require specifying a prior, which if not correctly specified would bias the results. Motivated by population genetics theory, we assume that the effect sizes follow a Gamma with unknown shape and scale parameters. Following that, it is straightforward to incorporate the distribution in a conditional likelihood and also to derive Bayes estimates. The methods we propose outperform existing methods and are robust to the assumed effect size distribution. Our methods are general in that they are developed for $P$-value and can be easily adapted for different effect size measures.
1779T  
Greater power for kernel-based tests using the likelihood ratio.  
eScience, Microsoft Research, Los Angeles, CA.  

Recently, tests for association between a phenotype and region- or gene-based sets of SNPs have attracted interest, in particular, for rare variants. One of the dominant approaches is to use a variance-component model, with a score test, such as implemented in SKAT. The score test in this setting can easily be computed in closed form, and has known asymptotic results, making computation of P values extremely efficient. In addition, because the score test does not require fitting of alternative model parameters, this test is much faster than a likelihood-ratio test (which does require such fitting). Finally, in light of the fact that asymptotic results for likelihood-ratio tests in this setting require unrealistic assumptions, one might conclude that the score test is the obvious method of choice. Surprisingly, however, we have found that there is good reason to use a likelihood ratio test over a score test in this setting. In particular, we have found that (1) the likelihood ratio test offers a substantial increase in power, and (2) that with just a handful of permutations, and a particular null distribution parametric form, that p-values can be computed reasonably efficiently (while controlling type 1 error). We systematically examined use of our approach on a number of real and simulated data sets spanning a wide range of important settings including case-control and continuous phenotypes, rare and common variants, and data confounded by relatedness and population structure. Overall, we found power to be greatly increased (about 50%) while still controlling type-1 error. To overcome the computational burden, we also describe a number of speedups for both tests that make genome-wide testing in the presence of population structure and relatedness feasible on data sets involving tens of thousands of samples, even for the likelihood-ratio test.

1780F  
Bayesian mixed model association statistics in linear time.  
P. Loh1,2, N. Patterson1, A.L. Price1,2.  
1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 3) Department of Biostatistics, Harvard School of Public Health, Boston, MA.  

Linear mixed models are a powerful statistical framework for identifying associated loci and avoiding confounding. Existing methods rely on the computation of a genetic relationship matrix, which requires time O(MN^2) (where N = #samples and M = #SNPs) and implicitly assumes an infinitesimal genetic architecture in which all variants are associated. Here, we propose a very fast O(MN) coordinate descent algorithm that can achieve the same results as existing methods, but can also be generalized to non-infinitesimal genetic architectures to increase power. In each iteration of the algorithm, the estimated effect size of each SNP is updated with its posterior mean given the estimated components of variance, the prior distribution on SNP effects, and the residual phenotype conditional on current estimates of all other SNP effect sizes. We define a retrospective score statistic computed from the phenotypic residuals upon convergence, which typically occurs in fewer than 10 iterations. When an infinitesimal (normal) prior distribution of effect sizes is used, the posterior mean effect sizes are best linear unbiased predictions (BLUPs) and optimize a ridge regression problem. Our iterative method thus calculates BLUP coefficients and residuals by applying coordinate descent on the vector of SNP effect sizes. In this case our statistic is equivalent to the GRAMMAR-Gamma mixed model association statistic (Svishcheva et al. 2012 Nat Genet); we observed a correlation of 1.000 in simulations. When modeling non-infinitesimal genetic architectures, our method achieves improved power as a consequence of more accurate estimation of sparse SNP effects: for example, in simulations in which only 5% of independent markers are causal (with the number of samples roughly equal to 10 times the number of causal markers), our approach achieves a 10% increase in test statistics at causal markers as compared to existing methods.

1781W  
Near equivalent calibration and power of joint and meta-analysis for association analysis of quantitative traits.  
C. Ma, T. Blackwell, M. Boehnke, L.J. Scott, the GoT2D investigators. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.  

In genome-wide association studies of quantitative traits (QTs), investigators typically analyze common genetic variants using linear regression within each study, and combine association results across studies using fixed-effects meta-analysis. For common variants (minor allele frequency [MAF] \( \geq \) 5%), linear regression is well-calibrated, and meta-analysis has near comparable power to joint analysis of the combined individual-level data (Lin and Zeng, 2010). In current sequencing and dense chip association studies, investigators wish to identify trait-associated low frequency variants (MAF < 5%), for which the calibration of linear regression analysis and the relative power of joint and meta-analysis has not been investigated. Here, we assess and compare the calibration and power of linear regression in joint and meta-analysis for QT association analysis across all allele frequencies, but in particular for low frequency variants, and examine the impact of sample size, additional covariates, and non-normally distributed QTs.

For common and low frequency variants (with minor allele count \( \geq 5 \)), simulation results show that: (1) linear regression-based joint analysis of normally-distributed QTs is well-calibrated; (2) sample-size weighted meta-analysis is well-calibrated and only slightly less powerful than joint analysis; and (3) inverse variance weighted meta-analysis is slightly anti-conservative. These results continue to hold when modeling the effects of additional covariates. For non-normally distributed QTs, joint and meta-analysis can become anti-conservative, but inverse-normal transformation of the QT remedies this problem while resulting in only modest loss of power. We are currently simulating quantitative trait data from the GoT2D sequencing study to assess the generalizability these simulation-based findings to real data.

1782T  
Estimating causal variant allele frequency, and thus efficacy of sequencing, at genetic loci identified by GWAS.  
P. F. O’Sullivan1, C.J. Hoggart2.  
1) Epidemiology & Biostatistics, Imperial College London, London, United Kingdom; 2) Genomics of Common Disease, Imperial College London, London United Kingdom.  

Genome-wide association studies (GWAS) have provided a large number of genotype-phenotype associations, but these can only be assumed to highlight genetic regions harbouring causal variants rather than identifying the causal variants themselves. However, sequencing of such susceptibility loci makes identifying the causal variants feasible, which is a crucial step in translating the findings from GWAS into medical application. Most genetic regions identified by GWAS have not been interrogated for the causal variants via sequencing, largely due to the significant cost involved, and when they have, this has tended to be on an ‘ad hoc’ basis according to the resources of individual studies or the significance of association signals at certain loci. The availability of the 1000G reference panel for imputation means that future GWAS will capture > 95% of common variants, so that subsequent sequencing is only worthwhile if the causal variant(s) is rare and has the power to be detected in the sample. Here, we introduce a novel method that estimates the posterior probability of the allele frequency of putative causal variants at a genetic locus, which can be used to determine whether sequencing is likely to yield causal variants or whether it is more likely that they are already present among the available genotyped or 1000G imputed SNPs. Our method exploits the SNP GWAS association P-values, their minor allele frequency and Linkage Disequilibrium (LD), to calculate the likelihood that the data are consistent with a causal variant(s) of a given effect size and frequency. While the probability computed is dependent on several modelling assumptions, we show how the method can be used to prioritise the sequencing of genetic loci with known effects on human traits or diseases based on GWAS results.
1783F
Testing Association without Calling Genotypes Allows for Systematic Differences in Read Depth and Sequencing Error Rate between Cases and Controls. G.A. Satten1, H.R. Johnston2, A.S. Allen3,4, Y.J. Hu2. 1) Division of Reproductive Health, CDC, Atlanta, GA; 2) Department of Biostatistics and Bioinformatics, Emory University, Atlanta GA; 3) Department of Biostatistics and Bioinformatics, Duke University, Durham NC; 4) Duke Clinical Research Institute, Durham NC.

Background: The quality of genotype calling for next-generation sequence data depends on read depth. Loci with high coverage can typically be called reliably, while those with low coverage may be difficult to call. In an association study, if case participants are sequenced to a greater depth than controls, the difference in genotype quality can introduce a systematic bias. This can easily occur when historical controls (e.g., data from The 1000 Genomes Project) are used as controls.

Methods: We propose directly comparing the proportion of calls for the minor allele between cases and controls, rather than comparing genotypes. We show how this proposal can be used to perform both single-marker test and gene-level test of rare variants (e.g., using modified versions of the burden test, variable threshold test or SKAT). We also show how this proposal can be extended to the situation where the per-call read error rate differs between cases and controls.

Results: Using simulated data and theoretical results, we demonstrate our proposals yield valid tests even in the presence of systematic differences in coverage rate between cases and controls, and show that in these situations, tests based on genotype have inflated size. We also show that power gains are possible using designs where we increase the number of controls while decreasing the read depth (while keeping total reads constant).

1784W

Meta-analysis of genome-wide association studies (GWAS) has led to the discoveries of common genetic variants for virtually every complex human disease. Recent advances in sequencing technologies have made it possible to extend association studies to rare variants. Because larger sample sizes are required to detect rare variants than common variants (with similar effect sizes), meta-analysis is as important to sequencing studies as to GWAS. Several research groups have recently developed meta-analysis methods for gene-level associations with rare variants under fixed-effect models. Those methods will lose power if the genetic effects are heterogeneous among participating studies. We propose novel multivariate random-effect models which allow the effects of multiple variants within a gene to vary among participating studies and develop the corresponding meta-analysis methods to perform gene-level association tests. Our methods take score statistics as input and thus can accommodate any study designs and any phenotypes. We produce the random-effect versions of all commonly used fixed-effect gene-level association tests, including burden tests, variable-threshold (VT) tests and sequence-kernel association tests (SKAT). We demonstrate through extensive simulation studies that the new methods are substantially more powerful than existing ones in the presence of moderate and high heterogeneity and maintain similar power when the heterogeneity is low. An application to the NHLBI Exome Sequencing Project data led to the discoveries of several genes associated with blood pressures that were previously undetected by fixed-effect methods. The relevant software is freely available.

1785T
A mixed model using both principal components and top markers corrects for population stratification and improves power. G. Tucker, B. Berger. Mathematics and CSAIL, MIT, Cambridge, MA.

In recent years, mixed models have been used extensively to calculate GWAS association statistics. Mixed models implicitly assume an infinitesimal genetic architecture, i.e., one in which all SNPs have a small nonzero effect; however, it is widely believed that disease phenotypes do not follow an infinitesimal model and that modeling a non-infinitesimal architecture would increase power. As a step in that direction, Lippert et al. (Sci Rep 2013) developed the state-of-the-art FaST-LMM Select method, which constructs the genetic relationship matrix (GRM) from a subset of the SNPs that are likely to be causal. FaST-LMM Select improves power over standard mixed models, especially when the phenotype is caused by few SNPs. However, limiting the GRM to a subset of the SNPs can result in insufficient correction for population stratification. As a solution to this problem, we propose a novel approach that pre-processes the data using principal components (PCs) and feeds the results to a program such as FaST-LMM Select, thereby leveraging the benefits of FaST-LMM Select, while properly correcting for population stratification.

Through simulation, we show that our framework corrects for population stratification without compromising power. When stratification exists, we observe that Fast-LMM Select inflates statistics on null SNPs ($\lambda_{GC} = 1.32\pm0.04$) whereas our method is properly calibrated ($\lambda_{GC} = 1.00\pm0.01$). Moreover, our method improves power over FaST-LMM Select (as measured by the mean $\chi^2$-statistic on causal SNPs: $11.4\pm0.1$ versus $10.6\pm0.2$), likely because the PCs reduce noise in selecting subsets of SNPs for the GRM. In addition, our method selects fewer SNPs to include in the GRM, yielding computational savings. In simulations without stratification, both methods perform nearly identically; thus our method retains the advantages of FaST-LMM Select over standard mixed models, such as improving power and correcting for confounding due to rare variants in spatially structured populations (Lippert et al., Sci Rep 2013; Listgarten et al., Nat Genet 2013).

1786F
Use of P-values to evaluate the probability of a genuine finding in a large-scale genetic association studies. O. Vsevolozhskaya1, CL. Kuo2, D. Zaykin3. 1) Mathematical Sciences, Montana State University, Bozeman, MT; 2) National Institute of Environmental Health Sciences, National Institutes of Health, USA.

To claim the existence of an association in modern genome-wide association studies (GWAS), a nominal P-value has to exceed a stringent Bonferroni-adjusted significance level. Despite strictness of the correction, a significant P-value does not indicate high probability that the claimed association is genuine. A simple Bayesian solution -- the False Positive Report Probability (FPRP) -- was previously proposed to convert the observed P-value to the corresponding probability of no true association. Although the FPRP solution is highly popular, it does not reflect probability that a particular finding is false. Here, we offer a simple POFIG method -- a Probability that a Finding is Genuine. POFIG enables one effectively to convert a P-value to the probability that a particular association with the trait is genuine. The validity of POFIG is supported by the results of a simulation study and the potential utility of our approach is discussed with reference to future GWAS discoveries.
A correction strategy for imputation across genotyping arrays. Y. Xie1, J. Rice2, L. Bierut3, R. Culverhouse4, N. Saccone2, E. Johnson2, D. Hancock4, COGEND collaborators. 1) Department of Psychiatry, Washington University in St. Louis, St. Louis, MO 63110; 2) Division of General Medical Sciences, Washington University in St. Louis, St. Louis, MO 63110; 3) Department of Genetics, Washington University in St. Louis, St. Louis, MO 63110; 4) Behavioral Health Epidemiology Program, RTI International, 3040 Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709-12194.

Genotype imputation methods are widely used to extend the utility of genome-wide association study (GWAS) data. Genotype imputation is a powerful approach because it can potentially identify causal SNPs that are untyped in the study data and can provide higher resolution for associated regions. However, bias can exist if different genotyping arrays are used and are unbalanced for case versus control subjects. Two currently used strategies of imputation across genotyping arrays are: (1) imputation based on the union of genotyped SNPs (i.e., SNPs on one of the two arrays); (2) imputation based on the intersection of genotyped SNPs (i.e., SNPs available on both arrays). Both of these strategies have weaknesses that are addressed by a third strategy we introduce here. We used data from chromosome 22 genotyped on Illumina 2.5M (32,903 SNPs) and 1M (14,071 SNPs) chips to illustrate properties of the three strategies. We demonstrate that the union strategy introduces bias (false positives) for SNPs available only on one array, presumably because the genotypes of some subjects are experimentally determined while others were imputed. The intersection strategy does not introduce this bias, but loses power for SNPs present only on one chip, due to the exclusion of the experimentally determined genotypes. We propose a hybrid approach that utilizes the genotypes for SNPs present only on one chip whenever the SNP has high imputation quality defined by the Imputation Quality Score (IQS) ≥0.9. The IQS is a score we previously introduced based on Cohen’s kappa rater agreement statistic. Using this approach led to good results with few spurious associations (false positives among 134,293 SNPs). In contrast, replacement of SNPs having high scores according to the IMPUTE2 internal quality metric gave rise to a higher proportion of spurious associations (126 false positives among 131,554 SNPs).

We compare the increase in power for our new approach to results from imputation based on the intersection of SNPs from the two arrays.

1787W

Testing Genetic Association at Untyped Rare Variants, an alternative to imputation based two-step approach. K. Ye1, Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA.

In recent years, we saw three important developments that empower us to investigate genetic association of rare variants (whose allele frequencies are below 5%) in genome-wide association studies. The first is the rise of large consortia, in which researches pool their samples together for meta-analysis. The second is the near completion of the 1000 Genome Project that provides detailed population genetic information on those rare variants and produces reference panels of haplotypes for a number of populations. The third, and perhaps most significant, is the availability of high-density genotyping arrays such as Beagle, IMPUTE and MACH, that use a reference panel of haplotypes to impute untyped variants based on genotyped markers. While imputation provides convenience and flexibility to test association at untyped variants, a large proportion of imputed variants, mostly rare variants, are often excluded in subsequent association test because of poor imputation quality.

In addition, the imputations are performed without considering disease status of subject, resulting biases at the risk alleles in disease populations, subsequently lowering statistical power of association studies. Among the imputed variants, we here present a novel association test that retains statistical powers especially at those rare variants with poor imputation quality. The new approach can be viewed as a special case of haplotype association tests, but it uses the information from haplotype reference panels, the same as those used for imputation. In short, for an untyped risk variant and a set of haplotypes at that locus, the expected frequencies of the haplotypes in a disease population are functions of (i) the relative risk of the risk allele, (ii) the proportion of risk allele in each haplotype. Therefore, once the proportion of risk allele on each haplotype is known, we can obtain a maximum likelihood estimate of the relative risk, and perform the likelihood ratio test for the association. Recursive tree models are applied on a reference panel to select such set of haplotypes at each untyped locus and to estimate the relative risk. We will illustrate how to apply this approach to the association tests on the imputed genotypes, this approach retains much higher statistical power at variants of low imputation quality. This is because, not only we use the information from a reference panel, the difference between disease population and normal population is also considered.

1789F

The more you test, the more you find: massive multiple testing does not promote spurious findings among top hits of association studies. D.V. Zaykin1, C.L. Kuo2, O.A. Vsevolozhskaya3, 1) Biostatistics Branch, NIH/NIH, Research Triangle Park, NC; 2) Department of Epidemiology and Biostatistics, Michigan State University.

High throughput whole-genome sequencing technology dramatically increased the multi-billion burden of association studies. High level of multiple testing is thought to promote false positives. To keep spurious findings at bay, it is common practice to require that findings should pass a significance threshold adjusted for the number of tests. As the number of tested variants now approached tens of millions, unrealistically large sample sizes would be needed for associations of modest magnitude to reach statistical significance. Paradoxically, in discovery studies where genetic variants are tested in an agnostic manner, “top hits” of a study become more likely to be true signals simply by performing many more tests, without any correction for multiple testing. We show by both statistical theory and simulations that when the rate of occurrence of true signals does not decrease as more tests are performed, findings with the smallest P-values are more likely to be genuine in studies with more tests. Our findings are supported by prior observations that GWAS studies with many tests enjoy a lower rate of false claims than traditional epidemiological and candidate gene association studies with fewer tests. This discrepancy has been attributed to the adoption of stringent significance thresholds and replication practices (Ioannidis et al., 2011). Replication is extremely important for eliminating erroneous claims due to study biases and errors unrelated to statistical chance. However, our results demonstrate that type-I error rates among best-ranking results of a study in fact diminish as more tests are performed. Our findings also demonstrate that even very high P-values may be a valid signal when large number of tests are performed. With regard to GWAS and sequencing studies, our findings suggest that for a given sample size, researchers should expect the top hits to be increasingly enriched with genuine associations as more variants are tested.

1790W

Efficient Algorithms for Multivariate Linear Mixed Models in Genome-wide Association Studies. X. Zhou4, M. Stephens1,2, 1) Human Genetics, University of Chicago, Chicago, IL; 2) Statistics, University of Chicago, Chicago, IL.

Multivariate linear mixed models (mLMMs) have been widely used in many areas of genetics, and have attracted considerable recent interest in genome-wide association studies (GWASs). However, existing methods for calculating the likelihood ratio test statistics in mLMMs are time consuming, and, without approximations, cannot be directly applied to analyze even two traits jointly in a typical-sized GWAS. Here, we present a novel algorithm for computing parameter estimates and test statistics (Likelihood ratio and Wald) in mLMMs that i) reduces iteration optimization complexity from cubic to linear in the number of samples; and ii) in GWAS analyses, reduces per-marker complexity from cubic to approximately quadratic (or linear if the relatedness matrix is of low rank) in the number of samples. The new method is effective EMMA (Efficient EMMA - Multivariate Association), the GEMMA (Genome-wide EMMA) algorithm to the multivariate case, making the likelihood ratio tests in GWASs with mLMM possible, for the first time, for tens of thousands of samples and a moderate number of phenotypes (<10). With real examples, we show that, as expected, the new method is orders of magnitude faster than competing methods in both variance component estimation in a single mLMM, and in GWAS applications. The method is implemented in the GEMMA software package, freely available at http://stephenslab.uchicago.edu/software.html.

1791T

Combined association and admixture mapping in Latinos. A. Brisbin1, A. Boyd1, S. Rachidi2, G. Lei3, A.C. Pereira4, J.E. Krieger4, 1) UW-Eau Claire, Eau Claire, WI; 2) Carleton College, Northfield, MN; 3) Health Sciences Institute, University of Sao Paulo Medical School, Sao Paulo, Brazil; 4) Biomedical Statistics and Informatics, Mayo Clinic, MN.

The use of ancestry information has the potential to enhance association mapping in admixed individuals, such as Latinos. Most existing methods of association and admixture burden of association tests have been developed for African Americans, who typically have ancestry from two populations. In this work, we explore a range of possible approaches for incorporating ancestry information from three populations into association mapping. We compare our methods on simulated data, and find that incorporating either genome-wide ancestry proportions or local ancestry provides a similar increase in performance over methods which ignore ancestry. Finally, we apply our methods to identify candidate loci for systolic blood pressure in Brazilians.
1792F

Expression quantitative trait loci (eQTLs) are genomic loci that regulate expression levels of genes. By assaying gene expression and genetic variation simultaneously on a genome-wide basis, scientists wish to discover groups of genomic loci that can affect the expression of a subset of genes. The problem can be viewed as a multivariate ‘reression’ with variable selection on both responses (gene expression) and covariates (genetic markers), including also multi-way interactions (epistasis effects) among covariates. Instead of learning a predictive model of expression levels given combinations of genetic markers, we start with an inverse modeling perspective. By conditioning on gene expression levels, we model the genetic markers via a sliced inverse model. An efficient dynamic programming algorithm is developed to determine the optimal slicing scheme. The inverse modeling approach can be effectively used for both independent screening and joint modeling of interactive genetic markers. Through simulation studies and real data examples in multiple tissues, we demonstrate how the proposed method achieves significantly improved power in detecting tissue-common and tissue-specific eQTLs compared to traditional approaches such as stepwise regression methods.

1793W
Admixed-MASTOR: Mixed-Model Association Mapping of Quantitative Traits in Genetically Admixed Samples with Related Individuals, T. Thornton, M. McPeek, 1) Biostatistics, University of Washington, Seattle, WA; 2) Statistics and Human Genetics, University of Chicago, Chicago, IL.

While genetic association studies for complex trait mapping have primarily focused on populations of European descent, more recent studies involve populations with admixed ancestry, such as African Americans and Hispanics. Genetic association studies in ancestral admixed populations offer exciting opportunities for identifying variants that underlie phenotypic diversity. At the same time, the heterogeneous genomic background and dependencies among sample individuals from admixed populations, including population structure and relatedness, pose special challenges for trait mapping. In these circumstances, it is necessary to devise statistical methods for association mapping that account for the diverse genomes of the sample individuals and are robust in the presence of a variety of complex sample structure settings. We propose ADMIXED-MASTOR, a mixed-model, retrospective score test for genetic association with a quantitative trait in the presence of ancestry admixture and relatedness. ADMIXED-MASTOR appropriately accounts for ancestry admixture by incorporating individual-specific allele frequencies in the mixed-model that are calculated on the basis of ancestry derived from whole-genome analysis. We demonstrate that ADMIXED-MASTOR can provide a substantial improvement over existing association methods, such as EIGENSTRAT (Price et al., 2007) and EMMAX (Kang et al., 2010), in terms of power and type 1 error in admixed samples with related individuals. We further demonstrate the utility of ADMIXED-MASTOR in an application to the majority cohort of more than 12,000 African Americans and Hispanics from the Women's Health Initiative study for the identification of genetic variants for a variety of clinical outcomes and quantitative traits.

1794T
Detecting local haplotype sharing and haplotype-phenotype association, Y. Guan, H. Xu1. 1) Pediatrics and Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Southeast University, Nanjing, China.

We present a statistical method to detect association between local haplotypes and phenotypes. The method relies on a two-layer hidden Markov model (HMM) developed previously to detect structure of local haplotypes. Briefly, we compute probabilities of (unspecified) haplotypes descending from a set of ancestral haplotype clusters, from which we quantify genetic distance between diploid individuals using local haplotype sharing -- the probabilities of two haplotypes descending from same haplotype clusters. We devise a novel model fitting method for the two-layer HMM, reducing the complexity from quadratic to linear (in the number of haplotype clusters) for diploid individuals; thus, it is feasible to apply our method to current genome-wide association studies (GWAS). We build a random effect model to link the genetic distance and phenotypes to test association. Because the genetic distance is quantified by local haplotype sharing, the associations detected are between phenotypes and (unspecified) haplotypes. Compare to traditional haplotype association methods, our method integrates out phasing uncertainty and avoids arbitrariness in specifying haplotypes. We recast the random effect model to a fix effect model through eigen decomposition of pairwise local haplotype sharing matrix; this allows us to take advantage of existing framework for association testing. We demonstrate the usefulness of our method by analyzing GWAS data sets from Wellcome Trust Case Control Consortium.

1795F
JEPEG: software for testing the joint effect on phenotype of eQTLs in a gene, D. Lee, V. Williamson, T. Bigdeli, V. Vladimirov, S. Bacanu, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA.

Expression of genes is known to affect many phenotypes/diseases. Given that gene expression is perturbed by nearby genetic variants commonly known as eQTLs, variants that are likely to affect many phenotypes in a multivariate manner. While modelling multivariately the action of eQTL/functional SNPs on phenotype is likely to increase detecting power for many traits, it is impeded by i) not measuring all functional SNPs in a gene and ii) the lack of access to individual genotypes. We overcome these obstacles by proposing JEPEG, a novel method/software testing for the joint effect of eQTLs in a gene by using only reported summary statistics. To achieve its objective, JEPEG i) directly imputes the summary statistics at unreported eQTLs and ii) tests for the joint effect of measured and imputed eQTLs. The direct imputation of summary statistics is achieved by employing the conditional expectation formula for multivariate normal variates. The joint testing is achieved by using summary statistics of measured and imputed functional SNPs and their genotype correlation matrix, as estimated from a relevant reference population (e.g.1000 Genomes). To decrease the number of degrees of freedom and, thus, increase power, before testing for the multivariate effect JEPEG first pools the genotypic information within functional categories using weights based on biologically relevant eQTL measures stored in a continuously updated database of human genes and eQTLs/functional SNPs affecting their expression. To decrease computational burden, the software implementing the method is written in C++ and is easily compilable to run on a variety of common operating systems such as Linux, Windows and Mac. We used JEPEG to analyze summary statistics from the Psychiatric GWAS Consortium (PGC) stage 1 (PGC1). Besides genes reported to harbor significant signals in PGC1, we detected multiple significant genes not harboring significant signals in PGC1, but subsequently reported to harbor significant findings in PGC stage 2. These results strongly suggest that JEPEG can improve the power to detect trait-associated genes by aggregating the functional information from a gene.

1796W
Detection Boundary and Higher Criticism Approach for Rare and Weak Genetic Effects, Z. Wu1, Y. Sun1, S. He1, J. Cho2, H. Zhao3, J. Jin4, 1) Mathematical Sciences, WPI, Worcester, MA; 2) Departments of Genetics and of Pediatrics, Yale, New Haven, CT; 3) Departments of Genetics and of Biostatistics, Yale, New Haven, CT; 4) Department of Statistics, CMU, Pittsburgh, PA.

Genome-wide association studies (GWAS) have identified many genetic factors underlying complex human traits. However, these factors only explained a small fraction of genetic contributions to these traits. It was argued that there are many more genetic factors remain undiscovered. These factors, each is individually weak at the population level, distribute sparsely across the genome. In this paper, we adapt the recent innovations on Tukey’s Higher Criticism to SNP-set analysis of GWAS and sequencing studies, and develop new theoretic framework in large-scale inference to assess the joint significances of such rare and weak effects for a quantitative trait. In the core of our theory is the so-called “detection boundary”, a curve in the two-dimensional phase space that quantifies the effect rarity and effect strength. Above the detection boundary, the overall effects of genetic factors are strong enough and allow for reliable detection. Below the detection boundary, the genetic factors are simply too rare and too weak for reliable detection by any statistical methods. We show that the HC-type methods are optimal in that they reliably yield detection once the parameters of the genetic effects fall above the detection boundary, and that many commonly used SNP-set methods are not optimal. The superior performance of the HC-type approach is demonstrated through simulations and the analysis of a GWAS data set of Crohn’s disease.
A novel method utilizing GWAS data identifies SNPs in the imprinted gene KCNK9 exhibiting parent-of-origin effects on BMI. C.J. Hoggart1, J.H. Zhao2, J. Luan3, F. Gomez4, G.B. Ehret5, P. Chasman2, S. Tajuddin6, A. Thomson7, N. Tşernikova8, T. Winkler9, W. Zhang4, B. Benyamin6,7,11, D. Evans12, S. Vedantam13, M.R. Jarvelin14,15,16, A. Scherag9, C. Riv reason7, I. Boreckij, I. Hischhorn17, R. Loos17, T. Frayling6, Z. Kutalik8, A. Hirschhorn18. 1) Department of Genetics of Complex Disease, Imperial College London, London, UK; 2) MRC Epidemiology Unit, Institute of Metabolic Science, Box, Addenbrooke’s Hospital, Hills Road, Cambridge, UK; 3) Division of Biostatistics and Statistical Genomics, Washington University School of Medicine, St. Louis, MO; 4) Geneva University Hospital, Geneva, Switzerland; 5) King’s College London, London, UK; 6) Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany; 7) Institute of Molecular and Cell Biology, Tartu, Estonia; 8) Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany; 9) Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK; 10) The University of Queensland, Queensland Brain Institute, St Lucia, Queensland, Australia; 11) Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 12) School of Social and Community Medicine, University of Bristol, UK; 13) Endocrine Division, Children’s Hospital, Harvard Medical School, MA; 14) Department of Epidemiology and Biostatistics, Imperial College, London, UK; 15) Institute of Health Sciences and Biocenter Oulu, University of Oulu, Finland; 16) Institut für Medizinische Informatik, Biometrie und Epidemiologie, Essen, Germany; 17) Department of Medical Genetics, University of Lausanne, Switzerland; 18) Broad Institute, Cambridge, MA; 19) Mount Sinai Hospital, New York, NY; 20) Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK; 21) University of Geneva, Geneva, Switzerland; 22) Swiss Institute of Bioinformatics, Lausanne, Switzerland.

It has been hypothesized that some genetic variants exert different effects on certain phenotypes depending on the parent from which they are inherited. Parent-of-origin effects may be caused by epigenetic factors such as methylation and histone modification which occur in genomic imprinting regions. Only a few family based linkage studies have explored whether SNPs exhibit parent of origin effects on traits such as blood pressure. Often, the findings of these studies remain un-replicated and the identified regions are often too large to provide meaningful interpretation. The only large-scale, genome-wide parent-of-origin study by Kong et al. 2009 focused primarily on type 2 diabetes and identified four loci; however, the discovery methodology required a combination of genealogy information and long-range phasing. Here, we present a novel method that is able to detect parent-of-origin effects using genome-wide genotype data of unrelated individuals. We demonstrate that parent-of-origin effects can be identified by increased phenotypic variance in the heterozygous (het) genotype group relative to that in the homozygous (hom) groups. Our rationale is that in the heterozygous group half of the population is mat-A/pat-B, and the other half is pat-A/mat-B, increasing the phenotypic variance in that group in the presence of a parent-of-origin effect. Sixteen GWAS cohorts participated in our discovery analysis totalling ~48,000 individuals. We tested all SNPs in known imprinted regions for parent-of-origin effects on BMI. One SNP reached significance adjusting for multiple testing (rs2471083 (1/C) variance (het vs. hom): 1.058 vs. 0.963, P = 1.07x10^-6). This SNP is located 100kb upstream of the gene KCNK9. Mutations in this potassium channel gene cause Birk-Barel syndrome. SNPs within 2kb have been shown to be associated with HDL cholesterol, adiponectin, and creatinine levels. We replicated the parent-of-origin association in four family-based studies; the combined analysis (of 3,016 heterozygous individuals) confirmed that individuals who carry the C allele paternally have 0.09 (SD unit) higher BMI on average than those carrying the maternal C allele (P=0.0031). Currently gene expression experiments are underway to investigate whether the variant may influence KCNK9 expression in a parent-of-origin fashion. Our method opens new avenues to exploit GWAS data of unrelated individuals in order to identify parent-of-origin effects.
equivalent to instrumental variable (IV) models employed in 'Mendelian randomization' between population. Since these are correlated via other effectors. (a) and (b) represent a variant that acts independently on one trait via its direct effect on an intermediate; since these are equivalent to instrumental variable (IV) models employed in Mendelian randomization. Our results offer an efficient method for IV model estimation in complex pedigrees as well as estimates of its Type I and Type II error rates.

1800T
Extended Mantel test for comparing differences in linkage disequilibrium between population. N. Tanaka. Biostatistics Section, National Center for Global Health and Medicine, Tokyo, Japan.

Checking the equality of population linkage disequilibrium (LD) structures is an important and often substantial in genetic association studies, especially if the target region includes disease susceptible genetic variants. This problem can be generalized as comparison of correlation structure between two populations. To solve this problem, the likelihood ratio test is commonly used especially in low-dimensional and regularity settings. In genetics, many researchers have been applying statistical procedures in which only elements of correlation matrices are used to calculate model-free test statistics, because the raw data which is used to estimate LD structures cannot be obtained in some cases. One of these is Mantel test (1967), which is originally proposed to compare distance or dissimilarity matrices for solving the problems inherent in explaining species-environment relationships in ecological studies. Mantel statistic is based on a simple cross-product term of each element of matrices, and the p-value is calculated based on the randomized distribution of ‘elements’. Thus this statistics represents not the relationship between populations but distance measures (matrices). However, Mantel test is still used even when investigators have the raw data to calculate correlation structure. Then, extended Mantel test statistic is proposed to compare two population structures, not the matrices. To address the variability of each element of matrices estimated in each population, the p-value for testing the null hypothesis, the correlation structure between two populations is equal, is calculated based on the sampling distribution of the weighted Mantel statistic via permutation of the observation. Proposed test is compared with Mantel test and the bootstrap based test which was proposed by Hrafni et al (2010). It is shown that this extended test statistic has broad utility and would derive a conclusion that agrees with that from other multivariate methods and intuitive insight from visualized method, such as Haplovieview, through several real examples.

1801F

Multivariate approaches to genetic analysis of complex diseases offer the promise of increased power to detect contributing genetic factors, as when affection status is combined with quantitative measures, or endophenotypes, of disease. In addition, analysis of endophenotypes may provide both a more focused search as well as insight into the physiological pathways and processes underlying disease. Bivariate joint association testing allows analysis of the (possibly) pleiotropic effect of a genetic variant while explicitly modeling the correlation between traits. However, (1) joint associations may be dominated by the univariate association of one of the correlated traits with the variant; (2) a test of joint association is often underpowered relative to univariate tests because it has two degrees of freedom; and (3) test results may pose problems of causal interpretation. Here we use simulation in SOLAR of genotypes and phenotypes on pedigree data to distinguish three possible causal relationships between a variant and two correlated traits A, B: (a) $V \rightarrow A \rightarrow B$; (b) $V \rightarrow B \rightarrow A$; (c) $V \rightarrow A$ and $V \rightarrow B$; where (c) represents a non-pleiotropic relationship between the variant and traits correlated via other effectors. (a) and (b) represent a variant that acts indirectly on one trait via its direct effect on an intermediate; since these are equivalent to instrumental variable (IV) models employed in ‘Mendelian randomization’, our results offer an efficient method for IV model estimation in complex pedigrees as well as estimates of its Type I and Type II error rates.

1802W
A Multi-Sample U-Statistic for Family-based Association Studies. M. Li1, Z. He2, D. Schaid3, M. Cleves4, T. Nick4, Q. Lu4. 1) Pediatrics, Univ Arkansas Medical Sciences, Little Rock, AR; 2) Biostatistics, University of Michigan, Ann Arbor, MI; 3) Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN; 4) Epidemiology and Biostatistics, Michigan State University, East Lansing, MI.

Family-based study design is one of the most commonly used study designs in genetic research. It has many ideal features, including being robust to population stratification (PS). With the advance of high-throughput technologies and ever-decreasing genotyping cost, the analysis of high-dimensional genetic data has become common for family studies. The yield from the analysis of these high-dimensional family-based genetic data can be enhanced by adopting computationally efficient and powerful statistical methods. In this article, we propose a general framework of family-based U-statistic, referred to as family-U, for family-based association studies. Similar to existing family-based methods, it can offer robust protection against PS. In the absence of PS, it can also utilize additional information (i.e., between-family information) for power improvement. The proposed family-U method makes no assumption of the underlying disease models, and is applicable for different phenotypes (e.g. binary and quantitative phenotypes), and various pedigree structures (e.g., nuclear families and extended pedigrees). Through simulations, we demonstrated that family-U attained higher power over a commonly used method, FBAT, under various disease scenarios. We further illustrated the new method with an application to a large-scale family data from Framingham Heart Study. By utilizing additional information (i.e., between-family information), family-U confirmed a previous association of CHRNAs with nicotine dependence.

1803T
Robust and powerful sibpair test for rare variant association. K. Lin1, S. Zöllner1,2. 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Psychiatry, University of Michigan, Ann Arbor, MI.

Background Modern sequencing technology allows investigating the impact of rare variant on complex disease. However, using a conventional case-control design, large samples are needed to achieve sufficient power for testing association of suspected loci to complex disease. In such large samples, population structure and batch effects can cause spurious signals. One approach to overcome stratification is family-based study design. Traditional methods such as the TDT and the FBAT are not especially powerful when applied to rare variants. Thus, there is a need to develop a method to efficiently use of family data to discover association between rare variants and disease phenotypes. Method We propose a novel test of a sample of affected sibpairs. In each sibpair, we estimate number of chromosomes shared IBD at the locus of interest. Based on this estimate, we assess the number of minor alleles located on shared chromosomes and the number of minor alleles on non-shared chromosomes. Shared and non-shared chromosomes are equally likely to contain variants that do not affect disease risk, while shared chromosomes are more likely to carry risk variants. Hence we test for an excess of shared variants on shared chromosomes. This test is robust to stratification as siblings have matched ancestries. Moreover, this design can correct for genotyping error and batch effects. As some regions are sequenced multiple times, sequencing error can be estimated by examining how often a sibpair who share two IBD chromosomes has inconsistent genotype calls and adjust the test to take into account the genotype error probabilities. We evaluate the power of this approach analytically as well as with computer simulations using a general model for the effect size of rare risk variants and considering different models of interaction between the locus of interest and the remaining genome. Result For most models with cumulative risk allele frequency <0.05, the proposed design shows superior power over the conventional case-control study given the same number of sequenced samples. Conclusion We introduce a new method for analyzing rare variants in families that is robust to population stratification and can incorporate other confounders such as genotyping error and batch effects. This new method is more powerful than a standard case-control design under wide-range of scenarios.
1804F
Analysis of Ultra High-Dimensional Data in Imaging Genetics Studies.
J. Kang, Y.J. Hu. Emory University, Atlanta, GA.

Many neurodegenerative diseases such as Alzheimer’s Disease (AD) are highly heritable and have been linked to brain regions of cognitive performance. This has spurred major interest in investigation of such diseases based on genetic and neuroimaging biomarkers. Recent advances in neuroimaging and genetics allow imaging genetics studies to collect both highly detailed brain images (>300,000 voxels) and genome-wide genotype information (>12 million variants). However, there is lack of statistical powerful methods and computationally efficient tools to analyze such very high-dimensional data, which has greatly hindered the impact of imaging genetics studies. In this work, we develop methods for genetic dissection of neuroimaging phenotypes. We build a statistical model that assesses the association between genetic variants and neuroimaging traits. Our method is tailored to the brain-wide, genome-wide association discovery, where both the entire genome and the entire brain are searched for dependence patterns. We also build a regression model that encompasses both genetic and imaging markers to predict the overall risk of the disease. A variable selection procedure is incorporated to search over the whole genome and the whole brain for variants that should be included in the regression model. We implement our methods in a user-friendly, computationally efficient, publicly available and well-maintained software package. We adopt the new computer technology, the General Purpose Computing on Graphical Processing Units (GPGPU), to make our analysis possible even on a personal computer (PC). We illustrate our method and software on simulation studies and analysis of the imaging genetics data from Alzheimer’s Disease Neuroimaging Initiative (ADNI).

1805W
A general regression framework for a secondary outcome in case-control studies.
S. Tchetgen Tchetgen. biostatistics, Harvard UBoston, MA.

Modern case-control genome-wide association studies typically involve in addition to the disease outcome and the genetic information, the collection of data on a large number of outcomes, often at considerable logistical and monetary expense. These data are of potentially great value to subsequent researchers, who, although not necessarily concerned with the disease that defined the case series in the original study, may want to use the available information for a regression analysis relating the genotype data to a secondary outcome. Because cases and controls are selected with unequal probability, regression analysis involving a secondary outcome generally must acknowledge the sampling design. In this work, the authors presents a new framework for the analysis of secondary outcomes in case-control studies. The approach is based on a careful re-parametrization of the conditional model for the secondary outcome given the case-control outcome and regression covariates, in terms of (a) the population regression of interest of the secondary outcome given covariates, and (b) the population regression of the case-control outcome on covariates. The error distribution for the secondary outcome given covariates and case-control status is otherwise unrestricted. For a continuous outcome, the approach sometimes reduces to extending model (a) by including a residual of (b) as a covariate. However, the framework is general in the sense that models (a) and (b) can take any functional form, and the methodology allows for an identity, log or logit link function for model (a). The approach is illustrated with an extensive simulation study and a genomewide association study of body mass index using data from a diabetes case-control study. The methods are found to be considerably more efficient than inverse-probability-weighting, and generally less biased than an estimator that stratifies or conditions on case-control status using a standard regression parametrization.

1806T
Mining the Human Phenome Using Allelic Scores that Index Biological Intermediates. D.M. Evans1, M.J. Brion1,2, L. Palenstom3, J.P. Kemp3, G. McManus2, J.L. Munafò3, N.J. Timpson1, B. St Pourcain3, A. Lawlor4,5, A. Dehghan1, J. Hirschhorn2,6, G.D. Smith2, 1) Soc Med, University Bristol, Bristol, United Kingdom; 2) Broad Institute, MIT and Harvard, Cambridge, Massachusetts; 3) School of Experimental Psychology, University of Bristol, Bristol, UK; 4) Department of Epidemiology, Erasmus Medical Centre, Rotterdam, The Netherlands; 5) 6Division of Genetics and Endocrinology and Program in Genomics, Children’s Hospital, Boston, Massachusetts; 6) Department of Genetics, Harvard Medical School, Boston, Massachusetts.

The common practice in genome-wide association studies (GWAS) and their meta-analyses to focus on the relationship between disease risk and genetic variants one marker at a time. When relevant genes are identified it is often possible to implicate biological intermediates and pathways likely to be involved in disease aetiology. However, single genetic variants typically explain small amounts of disease risk. Our idea is to construct allelic scores that explain greater proportions of the variance in biological intermediates, and subsequently use these scores to data mine GWAS. To investigate the approach’s properties, we indexed three biological intermediates where the results of large GWAS meta-analyses were available: body mass index, C-reactive protein and low density lipoprotein levels. We generated allelic scores in the Avon Longitudinal Study of Parents and Children, and in publicly available data from the first Wellcome Trust Case Control Consortium. We compared the explanatory ability of allelic scores in terms of their capacity to proxy for the intermediate of interest, and the extent to which they associated with disease. We found that allelic scores derived from known variants and allelic scores derived from hundreds of thousands of genotyped markers explained significant portions of the variance in biological intermediates of interest, and many of these scores showed expected correlations with disease. Power calculations confirm the feasibility of extending our strategy to the analysis of tens of thousands of molecular phenotypes in large genome-wide meta-analyses. We conclude that our method represents a simple way in which potentially tens of thousands of molecular phenotypes could be screened for causal relationships with disease without having to expensively measure these variables in individual disease collections.

1807F
Evolutionary Triangulation: A Novel Approach for Filtering Association Results.
M. Huang1, C. Amos2, L. Muglia3, S. Williams1. 1) Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, NH; 2) Department of Family and Community Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH; 3) Center for Prevention of Preterm Birth, Perinatal Institute, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH.

GWAS analyses have been successful in identifying many genetic associations with common complex diseases. Despite the successes, much of the heritability of complex disease remains unexplained. One possible explanation is that GWAS analyses suffer from an inflation of type II error rate due to overly conservative correction for multiple testing. Thus, it is important to develop methods that can minimize this type of error while still controlling type I error. Our hypothesis is that the distribution of genetic variants that associate with a disease will mirror the distribution of the disease among populations, and therefore, we can use the population genetic structure as a filter to reduce type II error. We propose a method to salvage real or likely associations in GWAS by applying evolutionary differentiation as a metric. This method, Evolutionary Triangulation (ET), requires populations with known disease prevalence data such that two have similar prevalence and the third is disparate. We use overlapping patterns of similarity and differences in FST to identify putatively associating genes. We tested ET by using three HapMap populations (YRI, CEU and GIH), because two are known to have similar prevalence for some cancers (e.g., in GIH and YRI melanoma and ovarian cancer are rare) but the third (CEU) is distinct and has a higher prevalence. Under stringent FST thresholds we first identified 33 ET SNPs by overlaps of high FST between CEU and GIH, high FST between CEU and YRI, and low FST between GIH and YRI. Then by testing all big SNPs back to genes, we identified 35 genes; five of which associated with melanoma, two with skin and hair color, and one with preterm birth. These diseases and traits are concordant with the discrepancy of phenotype distribution among these populations. This ability to identify genes with prior validated associations for disease provides a proof of principle that distribution of genetic variation can serve as an effective filter to find genes that associate with certain traits or diseases. Using a melanoma GWAS dataset and three previously mentioned three HapMap populations, we showed that an ET SNP subset was significantly enriched for GWAS ‘hits’ (p<0.05) compared to a subset of SNPs generated through random sampling, showcasing ET’s ability to identify putative GWAS associations. Our method, which is unbiased in terms of gene function, provides a potential means to address inflated type II error.
1808W
Detecting a weak association by testing its multiple perturbations.
M.T. Lo1,2, W.C. Lee1,2. 1) College of Public Health, Graduate Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan; 2) Research Center for Genes, Environment and Human Health, National Taiwan University, Taipei, Taiwan.

Genome-wide association studies (GWAS) have successfully identified several genetic variants associated with complex traits and diseases. These variants only account for a small proportion of heritability for traits and have miniscule effects. One therefore needs a study with a very large sample size (the number of subjects, $n$) to detect these weak associations. We refer the method of increasing sample size as the ‘n-based method’. The $n$ of a study can of course be increased but only to an extent. We propose a novel approach that aims for increasing the sample size in a different direction: the number of variables ($p$), which is denoted as the ‘p-based method’. For a putative genetic variant, the p-based method integrates the interactions (or more aptly, the ‘perturbations’) between the variant and other variables. In contrast to other dimension reduction methods, here we openly take advantage of the very large $p$. Theoretical power calculations and computer simulations show that this p-based method outperforms the traditional n-based method by a wide margin, when $p$ can be made very large, say, to the thousands or millions. The method is also applied to a real data from a GWAS of age-related macular degeneration (AMD). The method detects a number of novel genetic variants associated with AMD.

1809T
A Bayesian approach to Detect Differentially Methylated Loci with Next Generation Sequencing using Integrated Nested Laplace Approximation.
L. Shuang, R. Podolsky, D. Ryu, V. George, H. Xu. Georgia Regents University, Augusta, GA.

DNA methylation at CpG loci is an important biomedical process involved in many complex diseases including cancer. In recent years, the development of next-generation sequencing (NGS) has been yielding large amounts of sequencing data, which makes the NGS platforms useful for many applications. In this high-throughput sequencing approach, DNA samples are treated with bisulfate, which converts unmethylated cytosines to uracils and leaves methylated cytosines intact. The bisulfate-treated samples are then sequenced with NGS platforms. As a consequence, the NGS data include the counts of methylated molecules or unmethylated molecules at each CpG site for each individual. We introduce a robust Bayesian approach for differential methylation analysis for NGS data, using Integrated Nested Laplace Approximation (INLA) for deriving posterior distributions. We performed extensive simulations to compare our proposed method to existing alternate methods. The simulation results illustrate that our proposed approach can detect more true positive differential methylation sites and less false positives. Additionally, our approach allows us to compute approximations to the posterior marginal directly by using Integrated Nested Laplace Approximation (INLA). Compared to Markov Chain Monte Carlo (MCMC) for generating posterior samples, our approach using INLA is computationally faster, easier to implement and overcomes potential convergence issues associated with MCMC.

1810W
Methodology for mitochondrial phenome-wide association studies.
J. Hall1, W. Bush1, S. Pendergrass2, R. Goodloe1, J. Boston1, E. Farber-Eger1, D. Crawford1, S. Mitchell1. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Center for Systems Genomics, The Pennsylvania State University, University Park, PA.

In this pilot study, we evaluate methods for the analysis of mitochondrial variation on multiple phenotypic outcomes—HDL-C, LDL-C, triglyceride levels, total cholesterol, type 2 diabetes, systolic and diastolic blood pressure, and mean corpuscular hemoglobin. This is a novel application of a phenome-wide association study (PheWAS), a recent innovation in study design where phenotypically rich datasets are examined for relatively few genetic variables. These studies are ideal for detecting pleiotropic effects for susceptibility alleles, which data suggest are prominent for mitochondrial variants. Mitochondrial genetic variation, typically represented as haplogroups, has been associated to a wide variety of traits, though it is commonly ignored in genome-wide association studies. Given these associations, and the general biological implications of mitochondrial dysfunction, mitochondrial variants are prime candidates for PheWAS analysis. For this analysis, we accessed data from the Epidemiologic Architecture of Genes Linked to Environment (EAGLE) BioVU study, a subset of the Vanderbilt University biorepository consisting of ~15,000 DNA samples from non-European Americans representing African Americans (n=11,521), Hispanics (n=1,714), and Asians (n=1,122). Restricting analyses to African Americans only, we estimated the effect of 130 mitochondrial variants captured by the Illuma Metabochip. Using mixed linear models, we calculated the variance explained by mitochondrial genetic variation for each of these phenotypes. We compared models with only nuclear principal components (estimating genetic ancestry) to models containing both nuclear and mitochondrial principal components (representing maternal lineage), additionally adjusting for both age and sex. Addition of mitochondrial principal components did not significantly influence model fits, indicating that adjustment with nuclear principal components is sufficient. Our initial analysis suggests that 2–3% of variance in HDL-C and triglyceride levels may be explained by mitochondrial genetic variation, but that less of an effect is observed for the other phenotypes assessed. In future studies, we will refine the analysis approach and apply it to the full multi-ethnic dataset, for several hundred phenotypes.
1811W
Mechanistic Phenotypes: An aggregate phenotyping strategy to identify disease mechanisms using GWAS data. J.D. Mosley1, S.L. Van Driest2, E.K. Larkin1, P.E. Weeks1, J.S. Weete3, O.S. Wells4, J.H. Karmes1, L. Bastarache5, L.M. Olson5, C.A. McCarty6, J.A. Pacheco7, G.P. Jarvis8, E.B. Larson9, D.R. Crosslin10, I.J. Kullo11, G. Tromp12, H. Kuivaniemi17, D.J. Carey13, M.D. Ritchie14, R. Li14, J.C. Denny14, D.M. Roden1. 1) Department of Medicine, Vanderbilt University, Nashville, TN; 2) Department of Pediatrics, Vanderbilt University, Nashville, TN; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA; 4) Biomedical Informatics, Vanderbilt University, Nashville, TN; 5) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 6) Essentia Institute of Rural Health, Duluth, MN; 7) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 8) Department of Genome Sciences, University of Washington, Seattle, WA; 9) Group Health Research Institute, Seattle, WA; 10) Biomedical and Health Informatics, Seattle, WA; 11) Division of Cardiovascular Diseases, Mayo Clinic, Rochester MN; 12) The Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, PA; 13) Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA, USA; 14) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

A single mutation can alter homeostatic mechanisms and give rise to multiple complex diseases such as diabetes mellitus, asthma, and cancer. The assessment of the extent of disease sharing among diseases focused on single disease, and we test here the hypothesis that broad disease mechanisms can be identified by detecting associations between genes containing low minor allele frequency (MAF) SNPs and ‘mechanistic phenotypes’, comprised of collections of related diagnoses. Our overall approach is to rank associations between non-synonymous SNPs (nsSNPs) represented on GWAS platforms and mechanistic phenotypes ascertained from electronic medical records (EMRs), and then seek enrichment in functional ontologies across the top-ranked associations. We studied two mechanistic phenotypes: (1) thrombosis, evaluated in a population of 1,655 African Americans; (2) five groupings of cancer diagnoses were evaluated in 3,009 European Americans. We used a two-step analytic approach whereby nsSNPs were first sorted by the strength of their association with a phenotype. Then, we evaluated the top-ranking nsSNPs for each model, selected using p-value thresholds of 0.01 and 0.05, were tested for enrichment in functional ontologies. For the thrombosis phenotype, 452 nsSNPs met the initial selection criteria, and the top-ranked nsSNPs were enriched in ontologies related to blood coagulation (>10 fold enrichment; Fisher’s p<0.002). Enrichment was driven by the F5, P2RY12, PTAFR, F2RL2, HLA-DP1 and LEP genes. A total of 833 nsSNPs were evaluated for the tumorigenesis phenotypes. Ontological enrichment was only observed with the reverse genetics models. The enriched ontologies were related to DNA repair (9.9 fold; p=8.6x10^-6), FANCP, SLX4, FANCQ, XRCCL1, BCRA1, FANCA, CHD1L, and microtubule processes (5.1 fold; p=1.6x10^-4) (NIN, BCRA1, KIF25, DNAH3, 3C1R, SF1, DST). Mechanism-oriented phenotyping using collections of EMR-derived diagnoses can identify fundamental disease mechanisms.

1813F
Accounting for gene-by-sex interactions in genome-wide association studies using random effects meta-analysis. N. Furlotte1, E.Y. Kang1, B. Han2,3, J.W. Joo1, E. Fischhoff1,12, J.S. Witte1,12. 1) University of California Los Angeles, Department of Computer Science, Los Angeles CA; 2) Broad Institute, Cambridge MA; 3) University of California Los Angeles, Department of Human Genetics, Los Angeles, CA.

Many complex diseases are influenced by genetic variations in multiple genes and non-genetic factors. In order to find the association between SNPs and disease, an extension of genomic propensity score (eGPS) (Zhao et al., 2012) was used to control for bias due to both genetic and non-genetic factors. Pathway analysis, which identifies biological pathway associated with disease outcome, was also used here to improve the power of eGPS. We hypothesize that the type I error rate of this approach will be closer to its nominal value over a wide range of null conditions and that its power will be greater than the sum statistic and principle component analysis based on the simulation study reported here.

1814W
Increasing the Power of Association: POPFAM Combines Arbitrary Affected Families, Unrelated Cases, and Reference Samples. W. Stewart1,2, M. Monti1,2,1) Nationwide Children’s, Columbus, OH; 2) The Ohio State University; 3) University of Pavia.

We have extended our genetic association software: POPFAM so that in addition to case-parent triads and unrelated controls, the genotypes of large affected families, unrelated cases, and (optionally) reference individuals (e.g. HapMap samples) can easily be incorporated into a single, more powerful, test of association. Furthermore, in order to maintain POPFAM’s speed and flexibility, we have adapted our most recent release: POPFAM-v2.0 beta to a parallel computing environment. In this setting, workloads are distributed across multiple cores of a single machine and/or across multiple nodes of a cluster. Overall, by combining data across heterogeneous designs (e.g. ascertained affected family studies with variable pedigree structures, case-control studies with unequal sample sizes, and unphenotyped individuals), POPFAM-v2.0 beta gives researchers the ability to increase power for detecting genetic associations. Our software is distributed as a part of the larger package EAGLET, which is freely available from the worldwide consortium http://www.mathcs.emory.edu/~stewart/POPFAM/.

1815T
Modified Random Forest Algorithms For Analysis of Matched Case Control Data or Case-parent Trio Data. Q. Li, E. Y. Bailey-Wilson. Human Genome Research Institute, NIH, Baltimore, MD.

Random forests (RF) is a machine-learning method useful to detect complex interactions among random markers related to a disease trait based on case-control samples. We propose a new modification of the RF algorithm for matched case-control, or family based (trio) data analysis. RF is an ensemble method, which analyzes data and summarizes results using a large number of classification trees. During the procedure, each classification tree uses a proportion of samples and a subset of predictors. An R package, rpart, has functions implementing classification tree analysis and it can be modified to accommodate different study designs by substituting its functions of classification trees based on a novel criterion. For ease of implementation, our method utilizes the rpart package to conduct classification tree analysis on a subset of the samples and predictors. Then our ensemble code, also written in R, summarizes results from all trees. For matched case-control, or case-parent trio data, we sample the set of samples in a matched set, or matched case, pseudo-controls set) to be fit to each classification tree. Different classification criteria are also proposed to accommodate the matched study design. To evaluate our method, we simulated matched case-control, and case-parent trio data, and applied our method to select the top 1% most important predictors. The results are compared with other machine-learning methods applicable for matched case-control data, including RF++, MDR, and trio Logic Regression.
1816F
Exploiting network methodology for rare variant association analysis.
H. Fier1, J. Hecker1, D. Prokopenko1, P. Costa1, S. Won2,3, K. Ludwig4,5, E. Silverman4,5, M. Mattei8en4,5, M.M. Noethen4,5,6,8, E. Mangold4,5,6,8, C. Lange1,6,6,9. 1) Dept. of Genomic Mathematics, University of Bonn, Bonn, Germany; 2) Department of Applied Statistics, Chung-Ang University, Seoul, Korea; 3) The Research Center for Data Science, Chung-Ang University, Seoul, Korea; 4) Department of Genomics, Life & Brain Center, University of Bonn, Bonn; 5) Institute of Human Genetics, University of Bonn, Bonn; 6) Channing Laboratory, Brigham and Women’s Hospital, Boston; 7) Department of Biomedicine, Aarhus University, Denmark; 8) German Center for Neurodegenerative Diseases (DZNE), Bonn; 9) Department of Biostatistics, Harvard School of Public Health, Boston.

In this communication, we propose a new analysis approach for rare variant analysis that exploits network methodology. The approach is designed for large genomic regions, varying effect directions, etc. and enables the analysis of the relationships between the variants. The benefits of incorporating network methodology in genetic research has been shown by a number of current works, e.g. for prioritizing candidate disease genes (Akula et al. 2011, Kacprowski et al. 2013) or protein-protein networks (Gurney and Oliva 2012). First, we construct networks based on the aggregated allele counts of rare variants for cases and controls. Subsequently, we introduce a statistic that tests the centralization tendency of the derived networks. In a simulation study, we assess the power of the methodology and compare it with standard rare variant association approaches. We illustrate the practical applicability of the approach by an application to a sequencing study for nonsyndromic cleft lip with or without cleft palate.

1817W
Joint Association Analysis for Family-based Sequencing Data Using a Family-Genetic Random Field Method. Z. He1, M. Li1, M. Zhang1, X. Zhan1, Q. Lu1. 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR; 3) Epidemiology and Biostatistics, Michigan State University, East Lansing, MI. Advance in high-throughput sequencing technologies has revolutionized genetic study of complex diseases, especially the discovery of associated rare genetic variants. However, the high-dimensional feature of the sequencing data and the low frequency of the rare variants limit the power of detecting rare variants in population-based sequencing studies. As rare variants likely aggregate in pedigrees, family-based sequencing studies have substantial potential to enhance the power of detecting rare variants, remaining the advantage of being robust to population stratification (PS). In this paper, we propose a family-genetic random field method (F-GenRF) to test for association of rare and common variants in a region with complex diseases in family while adjusting for covariates. The new method is computationally efficient with following advantages: 1. It accommodates to a variety of disease phenotypes (e.g., quantitative, dichotomized or count phenotypes), and various pedigree structures (e.g., nuclear families and extended pedigrees); 2. It offers robust protection against PS using within-family information, and is able to utilize between-family information for power improvement in the absence of PS; 3. It is robust to the misspecification of within-family correlation structure. Through extensive simulations, we demonstrated these advantages by comparing the proposed F-GenRF with several existing association methods in family-based sequencing study.

1818T
Simulation study for rare variants approaches in family-based and case-control data. C. Herold1,2, H. Fier1, J. Hecker2, D. Prokopenko2, C. Lange1,3. 1) Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA; 2) German Center for Neurodegenerative Diseases (DZNE), D-53175 Bonn, Germany; 3) Department of Genomic Mathematics, University of Bonn, 53127, Germany.

After the great success of genome-wide association studies (GWAS) the current research focus now on the development of high-throughput sequencing technology to address the "missing heritability" problem. The majority of these high-throughput sequencing loci are rare variants i.e. loci with a minor allele frequency of less than 1% and therefore most of the single marker analyses for genetic association, e.g. Armitage-trend test, do not provide sufficient power [Lange and Laird, 2002]. For this reason, several new approaches for rare variants have been developed like the burden tests Combined Multivariate and Collapsing (CMC) test [Li and Leal, 2008], weighted sum approaches [Madsen and Browning, 2009; Price et al., 2010] and a replication-based strategy [Ionita-Laza et al., 2011] or non-burden tests like SKAT [Wu et al., 2011] and C-Alpha [Neale et al., 2011], but mostly for unrelated individuals. In our study we adopted these methods to family-based data, including a spatial approach which is taking in account the physical location of the variants [Fier et al., 2012; Bonetti and Pagan, 2005]. The advantages of family-based data are the robustness against population structure and they tend to be more powerful due to the increased likelihood of affected relatives to share the same rare disease variants [Laird and Lange, 2010]. Furthermore, the selection appropriate controls for rare variants in case-control studies can be a challenging task because of the very low frequencies. We will present the results of our simulation study for different rare variant approaches in family-based and case-control data which show that the power can be increased using family-based instead of case-control data.

1819F
Finding co-regulated transcripts associated to cooperating eSNPs. A. Kreimer1,2, I. Pe’er1. 1) Department of Biomedical Informatics, Columbia University, New York, USA; 2) Center of Computational Biology and Bioinformatics, Columbia University, New York, USA.

Cooperation of multiple SNPs that affect the same gene is a common feature of complex diseases. Such cooperations are often observed in case-control and family studies and in some cases can increase the power to detect rare variants. Finding these cooperators eSNPs by exhaustve search has proven to be statistically difficult. In this paper we utilizes sequencing data with transcriptional profiles in the same cohort to identify two kinds of usual suspects: eSNPs that alter the physical location of the variants [Fier et al., 2012; Bonetti and Pagan, 2005]. The advantages of family-based data are the robustness against population structure and they tend to be more powerful due to the increased likelihood of affected relatives to share the same rare disease variants [Laird and Lange, 2010]. Furthermore, the selection appropriate controls for rare variants in case-control studies can be a challenging task because of the very low frequencies. We will present the results of our simulation study for different rare variant approaches in family-based and case-control data which show that the power can be increased using family-based instead of case-control data.

1820W
Genome-wide association detection power increased by observation of cis-acting mRNA. A. Renwick1, J.W. Belmont2, C.A. Shaw1. 1) SCBMB, Baylor College of Medicine, Houston, TX; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Discovery of genetic loci causally associated with a trait of interest is hindered by the small effect typically attributable to any one locus. For alleles whose effect is mediated by a change in level of expression, mRNA offers an intermediate observation which closes the gap between allele status and trait. For such alleles, including mRNA data together with genotype and trait increases the power to detect true association.
1821T
Is the Tie-Corrected Mann-Whitney-Wilcoxon Test a Promising Alternative to the Cochran-Armitage Test in the Analysis of Genetic Association Studies? S. Welke1,2, A. T scholar1, 2) Dept of Biostatistics, CINN Mannheim/Univ of Heidelberg, Mannheim, Germany; 2) Dept of Medical Biostatistics, Epidemiology and Informatics, Univ of Mainz, Germany.

The most widely used procedure of testing for association between a SNP and disease status in genetic association studies is the Cochran-Armitage (CA) test. As argued by Welke & Ziegler (2012) (Hum Hered 73:14-17), a conceptually more appealing though asymptotically equivalent variant of the CA test is Wald’s maximum likelihood test about the regression coefficient associated with the SNP under analysis. Since none of these tests is uniformly most powerful, the question of possible competitors providing better power against certain alternatives arises. With a view to ease of implementation in the analysis of GWAS, a promising alternative choice is the classical Mann-Whitney-Wilcoxon (MWW) test corrected for the large amount of ties occurring in SNP data taking on only three possible values (0,1, or 2). The power of both the CA and the MWW test can be computed exactly for sample sizes of up to 50 per group with reasonable execution time of the programs. For larger sample sizes, one has to rely on Monte-Carlo simulation. The objective of this contribution is to identify alternatives against which the MWW test provides higher power as compared with the CA test.

1822F
1) Centre for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 2) Department of Urology, (SIUT), Karachi, Pakistan; 3) Department of Histopathology, (SIUT), Karachi, Pakistan; 4) Department of Human Genetics, University of Health Sciences (UHS), Lahore, Pakistan.

Abstract: Immune regulatory molecules are major players in renal allograft rejection. Among these, the chemokine, monocytes chemo-attractant protein 1 (MCP-1/CCL2), has been reported as an independent predictor for late renal allograft loss. In this study, an attempt to devise a non invasive marker, urinary MCP-1/CCL2 levels and their association with rejection episodes were determined in renal transplant patients from Pakistan. A total of 440 urine samples were selected for this study that include biopsy proven cases of rejection (187), non-rejection (48), No abnormality detected; NAD (174), and interstitial fibrosis and tubular atrophy; IFTA (47). Additionally, stable graft (42) and healthy controls (116) were also analyzed. The study was approved by the Institutional Ethical Review Committee and conformed to the Tenets of the Declaration of Helsinki. The quantification of urinary MCP-1/CCL2 levels was measured using the Human MCP-1/CCL2 Quantikine ELISA kit. The data were analyzed using the Statistical Package for Social Sciences (SPSS) and Med Calc softwares. The mean values of urinary MCP-1/CCL2 levels among rejection, IFTA, stable graft, NAD and healthy controls were 926.2±65.9 pg/mL, 306.5±87.7, 253.8±57.3, 29739.2 and 48.2±20 pg/mL respectively. MCP-1/CCL2 levels were significantly different between the rejection vs. controls (p=0.0001), rejection vs. stable graft (p=0.0001), rejection vs. NAD (p<0.0001) and rejection vs. IFTA (p=0.0001). Significant difference were also found among IFTA vs. control (p=0.0001) and IFTA vs. NAD (p=0.0024). The ROC curve analysis also showed significant differences among these groups. The area under curve (AUC) for rejection vs. control was 0.97. At a cutoff value 94.8 pg/mL, the sensitivity and specificity were 95% and 91%. However, in the rejection vs. NAD, rejection vs. stable graft and rejection vs. IFTA, the AUC were 0.76, 0.82 and 0.82 with a sensitivity of 55%, 82% and 78% a specificity of 85.4%, 74% and 79% respectively. The results showed increased levels of urinary MCP-1/CCL2 in the allograft rejection patients. This non-invasive investigation of MCP-1/CCL2 levels in urine, with others markers, may help in monitoring early rejection episodes.

1823W
A nonparametric Bayesian model for clustering time course gene expression profiles. D. Manandhar1,2, B. Engelhardt1,2,3.
1) Institute for Genome Sciences and Policy, Duke University, NC 27708, USA; 2) Dept of Biostatistics and Bioinformatics, Duke University, Durham, NC 27708, USA; 3) Department of Statistical Science, Duke University, Durham, NC 27708, USA.

Time course gene expression profiles have been used to understand numerous dynamic biological processes involving the cell-cycle, gene interactions, and genomic response to specific biochemical stimuli. In order to analyze these data, we have developed a nonparametric Bayesian mixture model that clusters the gene expression profiles using a Gaussian process (GP) to model gene expression trajectories. Numerous models -- such as mixed-effects model with B-splines, hidden Markov models, and models that treat gene expression values as samples from multivariate normal or t-distributions -- have been developed for time-course gene expression data. Each of these models has analytic limitations that our GP-based nonparametric Bayesian mixture model handles efficiently: a) we use a GP prior as the base distribution for Dirichlet process mixture model, where each time series is a continuous distribution instead of discrete time points; b) we let the number of clusters be determined by the data; c) the model is robust to missing or noisy time point measurements; d) GPs naturally capture equally spaced time interval experiments or unequal intervals. This model assumes gene expression levels at adjacent time points are correlated; moreover, we allow cluster-specific models of time-point correlation to capture the dynamic level of interdependence between time-points in different clusters. Each cluster of genes can thus be interpreted, identifying gene products that respond to a stimulus early and fast, and others that respond slower and with a delay. Our model investigates the degree to which whether a specific gene expression trajectory belongs to a cluster, and also implicitly estimates a similarity matrix between genes that can be used in downstream analyses, including gene network models. When we applied this model to gene expression profiles across human fibroblasts responding to the response of human fibroblasts to serum, we found dozens of distinct, interpretable gene expression patterns, versus ten clusters identified in the original analysis. We validate the robustness of patterns from the out-of-sample prediction rates, by looking for expression patterns that are biologically plausible to a cluster member and by analyzing the mechanistic interpretation of each cluster trajectory. The larger number of distinct patterns of gene expression response suggests additional modes of transcriptional regulation in serum-exposed fibroblast cells, which we aim to study further.

1824T
Enrichment of functional information (545 annotation tracks) in GWAS hits. S. Gagliano1,2, M. Barnes3, M. Weale1, J. Knight1,2,3.
1) Centre for Addiction and Mental Health, 250 College Street, Toronto, Ontario M5T 1R8 Canada; 2) Institute of Medical Science, University of Toronto, 1 King’s College Circle, Toronto, Ontario, MSS 1A8 Canada; 3) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, John Vane Science Centre, Charterhouse Square, London, EC1M 6BQ, UK; 4) Dept. of Medical & Molecular Genetics, King’s College London, Guy’s Hospital, London SE1 9RT, UK.

There are limits to genetic discovery based on association information alone. One thing the function of the variant in question can help determine whether a variant is non-synonymous, within a splice site, in a conserved region or in an enhancer region. Such information is being used to refine a methodology to identify SNPs that are likely to be causal based on their functional attributes, rather than relying solely on statistical results from GWAS.

There are limits to genetic discovery based on association information alone. One thing the function of the variant in question can help determine whether a variant is non-synonymous, within a splice site, in a conserved region or in an enhancer region. Such information is being used to refine a methodology to identify SNPs that are likely to be causal based on their functional attributes, rather than relying solely on statistical results from GWAS.
1825F
Use of complementary expression-based data provides improved gene-based prioritization of Crohn's disease associated loci. K. Ning, K. Gettler, J.H. Cho, NIDDK Inflammatory Bowel Disease Genetics Consortium. Digestive Diseases, Department of Internal Medicine, Yale University, New Haven, CT, USA.

Purpose: We have recently identified 144 Crohn's disease (CD) loci containing 1,404 candidate genes in a large case-control meta-analysis. Bioinformatic analyses and gene co-expression networks identified a striking overlap between CD loci and genes implicated in mycobacterial susceptibility and differentiation of macrophage subsets (M1, M2). We sought to improve prioritization of CD-associated genes by integrating GWAS results with a variety of gene features. Methods: We developed a reference list of genes with strong evidence of involvement in CD pathology and tested for enrichment of a variety of different gene features, including expression level in the intestine, disease-based differential gene expression, eQTL data, association with other auto-immune diseases and gene-level CD association p-value. We then built a logistic regression model to optimize weights of gene features for prioritizing CD-associated genes. Genes having strong evidence of involvement in CD pathology were used to label the dependent variable and all features under study were used as predictors. Step backward model selection was then carried out to keep the most important predictive features in the model. Results: In our final model, six features enriched in the reference genes were kept after model selection and their weights were estimated. After building the regression model, we assessed its performance regarding its statistical fitness and its capacity to provide biological insight: (1) Analyses of area under the model's receiver operator curve with permutations showed that our model fit the data well statistically (AUC = 0.8), (2) Comparison of top genes prioritized by our model and by GWAS data showed that, unlike genes implicated solely by association evidence, genes specifically prioritized by our model were more enriched in CD-relevant pathways, notably IFNγ signaling (FDR adjusted p-value=1e-7), and (3) Genes specifically prioritized in our model were significantly more likely to be differentially expressed between M1-M2 macrophage subsets, (which is modulated by IFNγ) compared with genes implicated by GWAS alone (p-value=2e-5). Taken together, our model successfully integrated GWAS and expression data to prioritize genes implicated in CD. We then developed a causal steps approach to mediation analysis at the gene/region level which enables expansion to simultaneously examine the joint role of both epigenetic and genetic variability in complex trait etiology, e.g. whether epigenetic mediation varies between both data types to determine if particular genetic regions are related to traits of interest. Second, it is of considerable interest how to leverage both data types to determine if particular genetic regions are related to traits of interest. Consequently, many large GWAS consortia are expanding to simultaneously examine the joint role of DNA methylation and genotype values for a particular gene, with correspondence suggestive of association. Similarly in methylation and genotype is found by constructing an optimally weighted average of the similarities in methylation and genotype values for a particular gene, with correspondence suggestive of association. Similarly in methylation and genotype is found by constructing an optimally weighted average of the similarities in methylation and genotype. For a significant gene/region, we then develop a causal steps approach to mediation analysis at the gene/region level which enables elucidation of the manner in which the different data types work, or do not work, together. We demonstrate through simulations and real data applications that our proposed testing approach often improves power to detect trait-associated genes, while protecting type I error, and that our mediation analysis framework can often correctly elucidate the mechanisms by which genetic and epigenetic variability influences traits. A key feature of our approach is that it falls within the kernel machine testing framework which allows for heterogeneity in effect sizes, nonlinear and interactive effects, and rapid p-value computation. Additionally, the approach can be easily applied to analysis of rare variants and sequencing studies.

1826W
The PhenX Toolkit: discovering and promoting opportunities for cross-study analyses. W. Huggins1, H. Pan1, D.M. Nettles2, J. Haines2, E. Eubanks2, T. Hendershot1, J.G. Pratt1, D. Maiese1, W.R. Harlan1, J. Haines2, H.A. Jenkins3, E.M. Ramos1, C.M. Hamilton1. 1) RTI International, Research Triangle Park, NC; 2) Retired, Associate Director for Disease Prevention, Office of the Director, National Institutes of Health, Bethesda, MD; 3) Center for Human Genome Research, Vanderbilt University, Nashville, TN; 4) National Human Genome Research Institute, Bethesda, MD.

The PhenX (consensus measures for Phenotypes and Exposures) Toolkit (https://www.phenxtoolkit.org/) is an online catalog of 339 measures of phenotypes and exposures for use in genetic and epidemiologic research. The PhenX Toolkit has 1,094 Registered Users and has been accessed by more than 100,000 visitors from 150 countries. Covering a broad scope of 21 research domains (Demographics, Cardiovascular, Psychosocial), PhenX measures can be used to prioritize and evaluate opportunities for cross-study collaboration. Investigators can find measures of interest by browsing or searching the Toolkit using the Smart Query Tool. For each measure, the Toolkit provides a description of the measure, the rationale for its inclusion, detailed protocol(s) for collecting the data, and supporting documentation. The Toolkit also provides custom data collection worksheets to support data collection and custom data dictionaries that support data submission to the database of Genotypes and Phenotypes (dbGaP). To support investigators who want to collect data via the Web, initial efforts have been made to develop Web-based versions of PhenX protocols. To promote cross-study collaborations, the Toolkit includes a new ‘Register Your Study’ feature allowing registered Users to browse basic information about each registered study (PI, research focus, number of subjects, study design), and a list of the PhenX measures related to each. PhenX concepts are included in the National Center for Biomedical Ontologies (NCBO) BioPortal. Using BioPortal tools, researchers can search for and identify relationships between PhenX terms and terms from over 200 ontologies, automatically annotate textual descriptions of data with PhenX concepts, and use PhenX terms to simultaneously search across 38 data repositories. PhenX measures are complementary to other ongoing standard measures initiatives and a comparison of available resources that provide researchers with recommended measures and protocols will be presented. Funding provided by NHLBI 5U01HG004597 and 3U01HG004597-03S3.

1827T
Kernel Machine Methods for Joint Testing and Integrative Analysis of Genome Wide Methylation and Genotyping Studies. N. Zhao, M. Wu, Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

Comprehensive understanding of complex trait etiology requires examination of multiple sources of genomic variability. Integrative analysis of these data sources promises elucidation of the biological processes underlying particular phenotypes. Consequently, many large GWAS consortia are expanding to simultaneously examine the joint role of DNA methylation and genotype. Two practical challenges have arisen for researchers interested in joint analysis of GWAS and methylation studies of the same subjects. First, it is unclear how to leverage both data types to determine if particular genetic regions are related to traits of interest. Second, it is of considerable interest to understand the relative roles of different sources of genomic variability in complex trait etiology, e.g. whether epigenetics mediates genetic effects, etc. Therefore, we propose to use the powerful kernel machine framework for first testing the cumulative effect of both epigenetic and genetic variability on a trait, and for subsequent mediation analysis to understand the mechanisms by which the genomic data types influence the trait. In particular, we develop an approach that works at the gene/region level (to allow for a common unit of analysis across data types). Then we compare pair-wise similarity in the trait values between individuals to pairwise similarity in methylation and genotype values for a particular gene, with correspondence suggestive of association. Similarly in methylation and genotype is found by constructing an optimally weighted average of the similarities in methylation and genotype. For a significant gene/region, we then develop a causal steps approach to mediation analysis at the gene/region level which enables elucidation of the manner in which the different data types work, or do not work, together. We demonstrate through simulations and real data applications that our proposed testing approach often improves power to detect trait-associated genes, while protecting type I error, and that our mediation analysis framework can often correctly elucidate the mechanisms by which genetic and epigenetic variability influences traits. A key feature of our approach is that it falls within the kernel machine testing framework which allows for heterogeneity in effect sizes, nonlinear and interactive effects, and rapid p-value computation. Additionally, the approach can be easily applied to analysis of rare variants and sequencing studies.
1828F Integrating Multiple Correlated Phenotypes for Genetic Association Analysis Through Heritability. J. Zhou1, M. Cho1,2,3, E. Silverman1,2,4, N. Laird1. 1) Harvard Sch Public Health, Boston, MA; 2) Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women’s Hospital, Boston, MA.

In genetic studies of complex diseases, many correlated disease variables may be analyzed in the hope of increasing power to detect causal genetic variants. A common statistical approach involves assessing the relationship between each phenotype and each single nucleotide polymorphism (SNP) individually and using a Bonferroni correction for the effective number of tests conducted. Alternatively, one can apply a multivariate regression or a dimension reduction technique, such as principal components analysis, and test the association with the principal components of the phenotypes rather than the individual phenotypes. Other previous approaches have developed methods for combining phenotypes to maximize heritability at individual SNPs. These approaches are not practical for population sample with genome-wide scans. In this paper, by taking advantage of the estimated heritability and co-heritability, we construct a maximally heritable phenotype which is a linear combination of the various phenotypes. Our approach estimates heritability globally and is therefore applicable to genome-wide scans. Theoretically, and through simulations, we compare our approach with commonly used methods and assess both the heritability of the overall phenotype and the power. Moreover we provide a guideline of how to choose the phenotypes for combination. Applications of our approach to a COPD genome-wide association study show the practical relevance.

1829W Analyses assessing enrichment of GWAS variants for non-coding annotations in the genome are upwardly biased. G. Trytka1,2, B. Han1,2, K. Slowikowski1,2, H. Xu3, X. Lu3, S. Raychaudhuri2,3,4. 1) Division of Genetics and Rheumatology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA, USA; 4) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

ENCODE and other efforts are annotating the non-coding genome, e.g. with DNase hypersensitivity (DHS) sites. Investigators now commonly assess whether these annotations might link trait-associated variants (e.g. from GWAS) to functional, non-coding regions. A ubiquitously applied approach is to quantify the overlap of GWAS SNPs (or SNPs in tight LD) with specific annotations, and assess if it exceeds chance. Standard practice is to assess statistical significance by comparison to randomly sampled SNPs, matched on minor allele frequency (MAF) or distance to transcription start site (dTSS). But, since GWAS is biased towards genomic regions with higher gene density and LD, standard enrichment analysis for annotations that cluster around genes might result in inflation estimates. To investigate this possibility, we compiled publicly available DHS regions from 217 samples collectively spanning 16.4% of the genome. Then to quantify type I error under different models, we simulated six different sets of GWAS catalogs, with all causal variants drawn from a single genome structure (promoter, coding, intronic, 5’ UTR, 3’ UTR, or intergenic). Of these, only promoters should overlap with DHS peaks. We then tested significance of enrichment by randomly drawing 1000 SNP sets matched for different parameters: 1) MAF, 2) dTSS, 3) number of LD partners (r2 > 0.8). Surprisingly, matching all SNPs only on MAF and dTSS resulted in p < 0.001 in 100% of instances. We observed that only by additionally matching on LD partners we were able to achieve appropriate type I error rates. As an alternative, we also defined null distributions by shifting DHS annotations locally by a random value between 5kb to +5kb. Using this approach, we observed close to the expected (5%) false positive rate at p < 0.05 within coding (9%), intronic (11%), 5’ UTR (4%), 3’ UTR (10%), and intergenic regions (7%); but we had high power to detect DHS enrichment for promoter catalogs (95%). Conclusion: SNP enrichment analyses can be upwardly biased to detect false associations. We present two strategies to address this: 1) match SNPs stringently on the dTSS, MAF and the number of LD partners or 2) shift annotations locally to define null distributions. It is critical that, as investigators assess whether trait-associated SNPs are enriched for specific annotations, these approaches are used to avoid inflated statistically significant results.

1830T dbVOR: An open source database system for managing phenotype and genotype information for complex trait studies. R.V. Baron1, Y.P. Conley2,3,4, M.B. Gorin1, 2,3,4, D.E. Weeks1,4. 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Department of Health Promotion and Development, School of Nursing, University of Pittsburgh, Pittsburgh, PA, USA; 3) Department of Ophthalmology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA; Jules Stein Eye Institute, Los Angeles, California, USA; 4) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

In genetic studies of complex traits, efficient data management is vitally important, yet often challenging. As technological advances permit genotyping at an ever increasing numbers of markers, handling such large scale data in a reasonable amount of time requires clever computational approaches. When sample size is large, or at multiple times, a consensus genotype for use during the analyses needs to be decided upon. Data cleaning will require samples or marker ranges to be discarded or re-measured. To address these issues, we have developed a portable open source database system, built using Python and MySQL, named dbVOR, after Vör, an inquiring Norse goddess of wisdom from whom nothing can be concealed. To handle large amounts of genome-wide genotype data, instead of storing genotypes individually, one by one, these data are split into chunks, which are then stored in blocks in the database. dbVOR also can store experiments done with small numbers of markers in a conventional way, an individual genotype at a time. To resolve multiple variants of samples, we generate agreement matrices displaying the genotype concordance, and either can output the consensus genotype if all genotypes agree, or the user can specify which genotype to trustworthily and to be used. dbVOR is run from the command line, controlled by a configuration file. dbVOR has facilities for importing data in Illumina and Affymetrix and tabular formats, for storing build-specific genetic and physical mapping and for handling multi-allelic markers, and parallelizing to resolve multiple different ID systems. dbVOR outputs in a variety of formats, including PLINK binary format and Mega2 annotated format. It also supports filtering of families, individuals, chromosomes, chromosome ranges, markers and alleles. As a desired subset of the data can easily be extracted. Our age-related maculopathy exome chip data has 247,519 SNPs stored in blocks on 1,058 people. It took 470 seconds (s) and 3.22 GB of memory on a 3.33 GHz Core 2 Mac to export 25,175 SNPs on chr. 1 to a binary PLINK file. There are 39 SNPs in both the blocked and conventional data: extracting from blocked only took 10 s, conventional 75 s, and extracting from both and resolving conflicts took 100 s. dbVOR is freely available from our http://watson.hgen.pitt.edu/register web site. This work was supported by NIH grants R01 EY008859 (PI: Gorin), R01 GM076677 (PI: Weeks), and ARRA supplement R01 EY009859-S14 (PI: Gorin).

1831F Mega2: enhanced data-handling for facilitating genetic linkage and association analyses. T. V. Baron1, Y. P. Conley2,3,4, M.B. Gorin1, 2,3,4, D.E. Weeks1,3,4. 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 3) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

It is common in a genetic study of a complex disease to use a variety of different analysis programs. Invariably, this will require reformatting the data into the precise input format required by each of the analysis programs used. Our Mega2 software seeks to facilitate this process by automating common data reformatting tasks, thereby markedly reducing the chance of data reformatting errors, making data setup much less error-prone, and streamlining the user the time of debugging, and maintaining data conversion scripts. We have recently extended the capabilities of Mega2 in a number of ways. In addition to accepting LINKAGE-format input files, it now accepts PLINK-format files, both text-based and compact binary formats, as input. Mega2 is also more efficient than it was previously, enabling it to handle large data sets faster. While continuing to support conversion to commonly used formats like Merlin, Mendel, SimWalk2, and SOLAR, Mega2 has now been extended to support data conversion to more analysis formats, including PLINK, Crandefolt, IOLDS, FBAT, MORGAN, BEAGLE, Eigenstrat, and Structure. For some output options, Mega2 supports the generation of high-quality plots of the results using our nplot R package, as well as generation of custom track files for use within the UCSC genome browser. It has also been extended to support bulk upload of data sets, saving the user the time of uploading, debugging, and maintaining data conversion scripts. It has recently been extended to allow batch processing of large number of analyses, saving the user the time of repeated data conversion steps.
1832W

The Research Program on Genes, Environment and Health (RPGEH) has assayed over 670,000 SNPs for 100,000 subjects in an effort to deliver a clean, consistent dataset for use for use in research on genetic influences on a broad variety of health conditions. The number of subjects created challenges for both calling genotypes and for ensuring the highest possible quality. To address these challenges, we created novel applications of machine learning to better assess sample and SNP assay quality. To determine SNP quality in the Axiom® genotyping assay, it is necessary to go beyond the confidence measures provided by the BRLMM-P genotyping engine. The gold standard for assessing SNP quality is manual inspection of fluorescent intensity profiles and genotype statistics of the population. The 100,000 samples were genotyped in 58 separate sets of samples that lead to more than 42 million intensity profiles and associated genotype inferences; manual assessment of these would take an impractically long time. To leverage the utility of manual assessment over all the genotyping, we created a support vector machine (SVM) classifier based on geometric and statistical measures of the intensity profiles, confidence distributions, and genotype statistics and trained it using manually graded SNPs. Optimization of the SVM hyperparameters was performed using cross-validation and an adaptive nullifier search. The resulting SVM classified SNP quality at a level comparable with manual grading and showed typical classification accuracy of 98-99%, with disagreements happening commonly in the frontiers of marginal quality scores. A significant number of genotype calls came from SNP locations in which double deletions (a copy number of zero) produced a null cluster not accounted for by BRLMM-P. The SVM classifier was adapted to detect the presence of null clusters and allow for recovery of these copy number variants. An SVM classifier was also created to aid in gender identification of assayed samples. Gender inference from the assay was invaluable in detecting sample and plate mix-ups. Because of wide variance in X and Y intensities and potential large chromosomal deletions, a simple threshold test is not sufficient. An SVM was trained against survey gender to look for problem misclassifications. Of more than 100,000 samples, several hundred misclassifications were detected and removed from the dataset.

1833T
FBAP: A pipeline for family-based quality control of pedigree structures and dense genetic marker data. A.Q. Nato1, N.H. Chapman1, C.Y.K. Cheung2, B. Brkenac1, E.M. Wijsman3,4. 1) Division of Medical Genetics, Department of Human Genetics, School of Medicine, University of Washington, Seattle, WA; 2) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 4) Department of Genome Sciences, University of Washington, Seattle, WA.

Inheritance vectors (IVs) specify the flow of founder alleles of each non-founder in a pedigree making them useful as a foundation for statistical genetic analyses, such as linkage analysis, association scans, and next-generation sequencing studies. Stringent quality control (QC) procedures for pedigrees ranging from small up to very large pedigrees and for genotype data of increasing resolution (multiallelic linkage marker panels, dense diallelic SNP marker panels, and high-throughput sequence data) are essential for accurate family-based analyses. We developed a family-based analysis pipeline (FBAP) that implements QC checks on genetic data of diverse breadth and depth, samples IVs based on validated pedigree structures and cleaned genotype data, and formats data for family-based analyses such as linkage analysis, association analysis, or imputation from sequence data. FBAP first identifies relationship/pedigree errors through the comparison of pairwise kinship coefficients based on pedigree structure vs. empirical kinship coefficient estimated using genotype data. Marker genetic locations (Kosambi) and physical positions are taken from the Rutgers Map and subsequently converted to their corresponding positions based on the Haldane function. An initial marker subpanel is selected from the dense panels by screening SNPs through LD-based SNP pruning, and by setting a minimum intermarker distance of 0.5 cm. Additional subpanels are obtained using differentially III-marked panels. The IVs of the initial subpanel are sampled by gl_auto of the MORGAN package conditional on validated pedigree structures, marker map, allele frequencies, and genotype data. Using each initial subpanel, Mendel-consistent errors of remaining markers that are not on the panel are detected with SGI, which treats the markers with possible errors. This process is performed on each subpanel to ultimately generate a final framework panel (composed of cleaned markers) and its corresponding sampled IVs. FBAP therefore allows researchers to select markers without Mendelian consistent or inconsistent errors, sample IVs based on validated data, and subsequently perform different types of family-based analyses.

1834F
Comparing a few SNP calling algorithms using low-coverage sequencing data. X. Yu1, S. Sun1,2. 1) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Cancer Center, Case Western Reserve University, Cleveland, OH.

Several single nucleotide polymorphism (SNP) calling programs have been developed to identify novel SNPs and mutations in next generation sequencing (NGS) data. However, low sequencing coverage presents challenges to accurate SNP calling. Moreover, commonly used SNP callers usually include several metrics for each potential SNP in their output files. These metrics are highly correlated in complex patterns, making it extremely difficult to select SNPs to do any further experimental validation. To compare the performance of SNP callers in a low coverage sequencing dataset, we first compare the SNP calling results generated from four algorithms, SOAPsnp, AtlasSNP2, samtools, and GATK, without any post-output filtering. We have a few findings. First, we find that SOAPsnp calls more SNPs than other algorithms since it has little internal filtering criteria. However, AtlasSNP2 reports the least number of SNPs since it has stringent internal filtering criteria. Second, using several cutoff values for the sequencing coverage of called SNPs, we find that filtering the SNPs with a higher coverage threshold improves the agreement among the four algorithms. Third, we explore the values of a few key metrics in each algorithm, and use them as post-output filtering criteria to maximize the agreement of SNP findings among algorithms. Our exploratory results show that high coverage regions or bases tend to have high calling qualities. We recommend the users to employ more than one SNP calling algorithm, and use coverage and calling quality as filtering criteria for reliable SNP identification.

1835W
Influence of low level contamination on variant calling and filtration of NGS data and its quality control. H. Ling, K. Hetrick, E. Pugh, J. Romm, K. Doheny. Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. In NGS labs, cross sample DNA contamination can be one of the major causes of genotyping errors and false positives. Contamination may be present in source DNA, introduced during library preparation or cross-talk during PCR amplification. To better understand its influence on data quality and QC data, we investigated 1) at what level variant calls start to be affected by contamination, 2) whether variant calling can be corrected by down-sampling alternate alleles to a fixed value or by providing an estimated contamination level, and 3) the upper limit of contamination at which the correction will still work properly to eliminate the influence. We chose a set of high quality samples and manually constructed contamination by proportionally merging BAM files from subjects with varying levels and types of combination (related/unrelated, ethnicities, 2-/3-way) at pre-determined level. Both contaminated and non-contaminated samples were called together by Unified Genotyper. A number of GC metrics (reproducibility, concordance and sensitivity to array data) were employed to evaluate the presence of contamination and quality of variant calls. VerifyBamID was used to estimate contamination levels to see how accurate the estimation is and the performance of down-sampling in correcting variant calls by feeding the estimated contamination level to caller. The contamination level was well estimated by VerifyBamID despite some variation across different types of contamination. Without any effort of correction, variant calls start to be affected when contamination reaches 3%. Down-sampling was effective in correcting moderate contamination when an accurate estimate of contamination level was provided during variant calling. However, when contamination level reached 10% or higher, down-sampling to remove contaminated variants did not work well and significant numbers of errors remained in the variant calls. Application and integration of VerifyBamID with variant calling allows us to detect contamination at early stages and significantly improves calling accuracy.

1836F
Several single nucleotide polymorphism (SNP) calling programs have been developed to identify novel SNPs and mutations in next generation sequencing (NGS) data. However, low sequencing coverage presents challenges to accurate SNP calling. Moreover, commonly used SNP callers usually include several metrics for each potential SNP in their output files. These metrics are highly correlated in complex patterns, making it extremely difficult to select SNPs to do any further experimental validation. To compare the performance of SNP callers in a low coverage sequencing dataset, we first compare the SNP calling results generated from four algorithms, SOAPsnp, AtlasSNP2, samtools, and GATK, without any post-output filtering. We have a few findings. First, we find that SOAPsnp calls more SNPs than other algorithms since it has little internal filtering criteria. However, AtlasSNP2 reports the least number of SNPs since it has stringent internal filtering criteria. Second, using several cutoff values for the sequencing coverage of called SNPs, we find that filtering the SNPs with a higher coverage threshold improves the agreement among the four algorithms. Third, we explore the values of a few key metrics in each algorithm, and use them as post-output filtering criteria to maximize the agreement of SNP findings among algorithms. Our exploratory results show that high coverage regions or bases tend to have high calling qualities. We recommend the users to employ more than one SNP calling algorithm, and use coverage and calling quality as filtering criteria for reliable SNP identification.
1836T Handling dimensional-genetics data using multilevel dimensionality reduction algorithms in genetic association studies. K. Cho1, 2, D.R. Gagnon1, 3, H. Wu1, 4, 1) Massachusetts Veterans Epidemiology Research and Information Center; VA Boston Healthcare System, Boston, MA; 2) Division of Aging, Department of Medicine, Brigham and Women’s Hospital/Harvard Medical School, Boston, MA; 3) Boston University School of Public Health, Boston, MA; 4) Computer Science and Networking, Wentworth Institute of Technology; Boston, MA.

In recent years with advancing genotyping and sequencing technologies, there are more genetic data available than one can thoroughly analyze with existing tools and methods. This has become one of the analytical challenges in the initial phases of big scale genetic studies. In whole-genome or genomewide exploration of novel associations, without a priori knowledge of specific regions or genes of interest, researchers rely on the available techniques and methods that perform association tests based on a wide sum of characterizing or assuming independence. In addition, traditional statistical procedures present eminent challenges in using these data, where the number of parameters is scalable larger than number of observations. These limitations may compromise the level of false positive results and power to conduct studies with acceptable level of confidence. One of the ways to reduce the impact of these limitations is to first perform a data dimension reduction process and then evaluate the resulting panels of markers. We propose several algorithms using traditional PCA, PCA incorporating heritability, LASSO, and combinations of these to evaluate the impact of type one error. Using the Genetic Analysis Workshop 18 simulated data, we apply the proposed algorithms on selected null regions of chromosome 3 with regards to the diastolic blood pressure phenotype. Then we use SKAT as the evaluation tool to perform association analysis using the resulting panels from each algorithm. We then systematically compare type I error rates among different algorithms with respect to the baseline panel. Our preliminary work using a subset of 2000 SNPs (1723 polymorphic SNPs) in unrelated samples shows a substantial reduction of the markers through applying proposed algorithms. Among these, the PCA-only algorithm resulted in 47 components comprised of 363 unique SNPs and PCA followed by LASSO resulted in 110 SNPs (32 components) left in the resulting panels, respectively. We are also investigating PCA incorporating heritability approaches using the family datasets. As the amount and dimensionality of the network of genomics data is only escalating in the coming years, a timely, relevant and practical analytical pipeline through multilevel dimensionality reduction techniques provides an efficient approach for the initial screening tool.

1837F Genotype imputation in the era of next-generation sequencing. G. Piatra1,2, E. Porcu1,2,3, C. Sidore1,2, F. Danjou1,2, M. Stern2, A. Mulas3, M. Zichichi4, F. Busonero1, F. Reiner1, R. Atzeni4, M. Lobing1, R. Pilu4, M. Marcelli5, B. Tarriere6, H.M. Kang7, A. Angius3,4, C.M. Jones1, D. Schlesinger1, F. Cucca2,3, G. Abeasis1, S. Sanna1, 1) University of Michigan, School of Public Health, Ann Arbor, MI, 48109; 2) Università degli Studi di Sassari, Dipartimento Scienze Biomediche, Sassari, 07100, Italy; 3) Istituto di Ricerca Genetica e Biomedica-CNR, Monserrato (CA), 09042, Italy; 4) Center for Advanced Studies, Research and Development in Sardinia - CR54, Pula, Italy; 5) DNA Sequencing Core, University of Michigan, Ann Arbor, MI, USA; 6) Laboratori di Genetica, Istituto di Ricerca Genetica e Biomedica-CNR, Monserrato (CA), 09042, Italy.

Genotype imputation is an essential tool to infer missing and untyped genotypes in genetic studies. Nevertheless, the performance of genotype imputation on rare (MAF <0.5%) and low frequency variants (MAF <1%) is still an unexplored field, and several issues must be specifically addressed: choice of reference panel, quality of input genotypes/haplotypes and imputation quality for less common variants. Recently, we completed whole genome sequencing of 2,120 Sardinians using a low-pass approach (average 4x coverage) as well as genotyping of 6,600 individuals enrolled in the SardiNIA project, using four different illumina Beadchip arrays: OmniExpress, Cardio-MetaboChip, ImmChop and ExomeChip. We used combinations of the first three arrays to perform genotype imputation using Sardiniq sequencing data (SardiSeq) as population-specific reference panel, and used the fourth array to evaluate accuracy as squared Pearson correlation R2 between dosages and the real genotypes. We also performed imputation using two 1000 Genomes Project (1000G) reference panels (ALL and EUR datasets). We observed that the use of SardSeq boosted imputation accuracy compared to the other reference panels, not only for genomewide distances, but also at the intragenic reference panels. Furthermore, we found that the imputation accuracy using the SardSeq panel on Europeans was on average similar to that reached using 1000G panels on the Sardinian population, for any combination of the same arrays, while for the African Panel a combination of SardSeq and 1000G haplotypes was more beneficial in Europeans than in Sardinians. Our results suggest that using a population-specific reference panel rather than public references (especially for different ancestral or geographically distant populations) yielded better imputation quality and, particularly for less common imputed variants, might reduce false positive signals in association analyses.

1838W Impact of Quality Control on the Heritability Analyses for Qualitative Traits. J. Liu1,2,3, T. Hoffmann1,4, E. Jorgenson5, J. Witte1,2,3,4, 1) Department of Epidemiology & Biostatistics; 2) Department of Urology; 3) Institute for Human Genetics; 4) Diller Family Cancer Center, University of California, San Francisco; 5) Kaiser Permanente Northern California Division of Research, Oakland, CA.

While a single SNP accounts for only a small fraction of the genetic variation in complex traits, recent work has shown that a substantial heritability of complex quantitative traits can be explained by considering all SNPs simultaneously from genome-wide association studies. This approach has been extended for evaluating the chip heritability of qualitative disease traits. However, analyzing disease traits requires more rigorous quality control than for quantitative heritability—or for a conventional GWAS—because slight case-control differences can be magnified into false evidence of heritability. In this study, we investigate various factors that may affect the assessment of disease heritability, and give a detailed quality control framework for accurately estimating heritability. Using two GWAS of prostate cancer, we show that failing to meet key quality control steps can vastly over- and underestimate heritability. In a GWAS of European Americans, the naïve prostate cancer heritability is an unreasonable 68%, whereas the ‘cleaned’ heritability is 59%. In a GWAS of African Americans, the naïve heritability is 22%, and the cleaned heritability is 32%. The key factors that lead to over and underestimates are (1) SNPs fail Hardy-Weinberg Equilibrium test and (2) individuals with genome-wide similarities greater than 0.025. Both of the ‘cleaned’ estimates are in line with what one would expect from previous heritability estimates for prostate cancer. These findings emphasize the need to use more stringent quality control criteria for evaluating chip-heritability in disease traits than one might expect based on heritability for quantitative traits or GWAS.
1839T
Visualization software for the efficient review of alternate genotype calls. Z. Xu, N. Pankratz. University of Minnesota, Minneapolis, MN.

The Illumina HumanExome BeadChip (the ‘exome chip’) has unique challenges over traditional GWAS arrays, as rare variants are often not clustered well using the default GenomeStudio GenCall/GenCall 2 algorithms. New algorithms have been developed to address this, such as zCall and optiCall, however no known software can visually compare genotype calls from different algorithms overlaid on raw intensity data in an integrated and efficient manner. We have developed platform-independent Java software to plot a marker's probe intensities for all individuals and to color-code any user-defined variable (e.g., gender, study, DNA source, genotyping batch, in addition to the genotype itself) thus allowing investigators to look for batch effects, detect if a SNP association is due to artifact, or to see if rare variants are being called accurately. We have also implemented an annotation system whereby the user can tag a marker (e.g., ‘unusable’, ‘copy number variant’, ‘pseudo-autosomal chrY marker’, ‘extra heterozygote clusters’, ‘ok’), using a single key stroke or mouse click which can also auto-advance to the next questionarable marker. Lists of questionable markers can be generated automatically using over a dozen criteria, including low call rate, excess low heterozygosity, mean theta values deviating from expected, etc.

In addition, alternate genotype calls can be imported and customizable colors/symbols can be used to denote points that are discordant for the various calling algorithms (e.g., samples called as missing by GenCall and as heterozygotes by zCall could be orange triangles). The user can then use the software to draw a square around these points and set them to what the user believes to be the proper genotype call. Once this is done for a series of markers, the reclustered data can be exported into the user's preferred format (e.g., GenCall). The ability to plot disparate genotype calls and then recluster them is not known to be available in any existing software, commercial or otherwise. This software is publicly available as free and open source and will allow researchers to triage, annotate, and report markers from any array as quickly and as efficiently as possible.

1840F
PhenoMan: Phenotypic data exploration, selection, management and quality control for association studies of rare and common variants. B. Li, G. Wang, S.M. Leal. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Next generation sequencing and other high-throughput technology advances have promoted great interest in detecting associations between complex traits and genetic variants. Phenotype selection, quality control (QC) and control of confounders are crucial and can have a great impact on the ability to detect associations. Although there are available programs to perform association analyses, e.g., PLINK, they cannot be used for comprehensive management and QC of phenotype data. To address this need PhenoMan was developed to 1.) select individuals based upon multiple phenotype criteria, 2.) replace missing covariate data, 3.) remove duplicates and individuals who are related, belong to another population or have incorrect sex specification, 4.) recode primary traits and covariates, 5.) transform and visualize quantitative traits, 6.) remove or windsorize outliers, 7.) select covariates for analysis, 8.) create residuals and 9.) provide summary statistics. A report is generated for each data set in order that the same protocol can be used to ensure consistency and harmonization between analyses: PhenoMan is a user friendly interactive program that integrates data exploration, management and QC using a unified platform. Proper QC of phenotypes before proceeding to association analyses is critical to ensure control of type I and II errors, reliable effect estimates and consistent results between studies. PhenoMan is highly beneficial for preparation of case, control and quantitative trait data for association studies using new data sets as well as those obtained from public repositories. The PhenoMan program and documentation are available at https://code.google.com/phroman.

1841W
Forensic Inference via a Genome Parade. Y. Chen1,2. R. Xia1,2. F.A. San Lucas1,2. S. Vattathil3,2. P. Scheet1,2,3,4. 1) Division of Biostatistics, School of Public Health, The University of Texas Health Science Center at Houston, TX; 2) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX; 3) Program in Human and Molecular Genetics, The University of Texas at Houston Graduate School of Biomedical Sciences; 4) Program in Biomathematics and Biostatistics, The University of Texas at Houston Graduate School of Biomedical Sciences.

Contamination of DNA may present problems or clues in studies using high-throughput next-generation sequencing data. Methods such as ContEst (Cibulskis et al, 2011, Bioint., 27:2801) and ContaminationDetection (Jun et al, 2012, AJHG, 91:839) have been shown to provide accurate estimates of contamination levels using complementary microarray data to look for deviations from expected allele-specific read counts. Here we present a statistical model that incorporates the dependence of these population allele frequencies, or linkage disequilibrium (LD), to improve component identification when encountering mixtures of two diploid genomes in next-generation sequencing data. Our method, which we call Genome Parade, uses a hidden Markov model similar to fastPHASE (Scheet & Stephens, 2006, AJHG, 78:629) to accommodate dependence of alleles potentially from a contaminator’s DNA, essentially presenting pairs of haplotypes ‘in turn’ from a model for genetic variation in front of the observed data for comparison (analogous to an ‘identity parade’). Using simulated and real data, we explore certain forensic applications, as existing methods offer limited power to identify the genotypes of a contaminating individual. For example, compared with an approach that ignores LD, our method increases the squared correlation between estimated genotypes and true genotypes of the contaminating individual from 57 to 82 in a scenario of 2% contamination and sequencing depth of 100. This is essentially an application of low-coverage LD-based imputation in the unusually difficult setting where we are interested in the rare source of DNA in a mixture. This may be useful when attempting to identify a contaminating individual in a subtle mixture of DNA, when plausible candidates are not readily available, or to aid in reducing a search space for brute-force methods that use large databases. In a related application, Genome Parade may be useful when indeed there exists a suspected source of DNA but when the alternatives do not come from an identifiable list. In such cases, the probability of the data may be easily calculated for the suspected source, but to quantify this evidence a comparison must be made to haplotypes drawn from the general population. Finally, incorporation of LD may aid in quantifying contamination in applications of targeted sequencing, since ignoring the dependence in the data indicates a false sense of precision when data are limited.
1842T

Discovery and replication of genetic interactions for quantitative lipid traits. E.R. Holzinger1, M. Farrall2, F. Drenos3, C.B. Moore4, I.B.C. Lipid Working Group4, S. Juvik1, H. Watkins5, F.W. Asselbergs6, B.J. Keating7, M.D. Ritchie1. 1) National Institutes of Health, National Human Genome Research Institute, Inherited Disease Research Branch, Baltimore, MD; 2) Center for Applied Genomics, Children’s Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA; 3) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 4) Center for Systems Genomics, The Pennsylvania State University, University Park, PA 16870; 5) The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6) Department of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA; 7) Department of Computer Science, University of California, Los Angeles, CA, USA; 8) Division of Genetics, Brigham & Women’s Hospital, Harvard Medical School, Boston, MA, USA; 9) Partners HealthCare Center for Personalized Medicine, Boston, MA, USA; 10) Department of Computer Science, University of California, Los Angeles, CA, USA; 11) Department of Human Genetics, University of California, Los Angeles, CA, USA.

The genetic etiology of human lipid quantitative traits may be further elucidated by considering interactions between variants. In this study, we perform a genome-wide interaction study (GWIS) for four different lipid traits - low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol (TC), and triglycerides (TG). Our analysis consisted of a discovery phase using individuals from five different cohorts (ARIC, CARIDA, CHS, FHS, and MESA; n~14,000) and a replication phase in a subset of the PROCARDIS cohort (n~6,400). Currently, there is no gold standard for performing a powerful GWIS. Due to the computational and multiple testing burden of exhaustively testing all pairwise interactions, filters are often applied before interaction testing. In this analysis, we used two different filters: 1. Main effects filter (MEF) - exhaustively test for pairwise effects in the replication phase that pass a main effect threshold of p<0.001. 2. Biological filter (BF) - SNP-SNP models generated with prior biological evidence supporting a potential interaction using Biofilter 2.0. We tested interactions using a linear regression model with a multiplicative interaction term for the two model SNPs. We selected SNP pairs from the discovery analysis with interaction p<0.001 and identified all SNP-SNP models that represent the same signal based on LD. We then tested these models in the replication dataset. We corrected for the number of interaction signals tested in the replication phase using a Bonferroni correction (MEF: 0.0003; BF: p<0.002). In the MEF analysis, 15 interaction signals replicated for HDL-C, 4 for LDL-C, and 12 for TG. No models replicated for TC. For the BF analysis, the TG trait analysis resulted in 4 interaction signals that replicated. No interactions replicated for the other lipid traits. Interestingly, the same SNP-SNP interaction replicated in both the MEF and BF analyses for TG (rs1263173 and rs12225230; p=4.8×10−4). Additionally, applying a multivariate model results in increased degrees of freedom and low statistical power. In this paper, we propose a meta-analytic approach based on a random effects model to identify loci involved in gene-environment interactions that pass a meta-effect threshold of p<0.001. The proposed approach is quite versatile, and can be used to study interactions in varying environmental conditions. Our approach is motivated by the observation that methods for discovering gene-environment interactions are closely related to random effect models for meta-analysis. We show that the interactions can be interpreted as heterogeneity and can be detected without utilizing the traditional uni- or multi-variate models for discovery of gene-environment interactions. We apply our method to combine 17 mouse studies containing a total of 4,965 animals. Using these studies, we identify 23 significant loci involved in HDL cholesterol levels. Among these, 23 loci are previously confirmed to have an effect on HDL cholesterol or closely related lipid levels in the blood, while 3 loci are novel. Several of these loci show significant evidence of gene-environment interactions in different environmental conditions. This approach focuses on a focus on a locus on chromosome 8 showing strong evidence of sex-by-environment interaction. Future work will consider interactions involving multiple environmental conditions.

1843F

Association analysis of gene-environment interactions in lipid profile using exome sequence data. Z. He, G. Wang, S. Leat, Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Although several modifications and extensions of MDR have in part addressed the practical problems, they are still limited in statistical analyses of diverse phenotypes, multivariate phenotypes and correlated observations. Furthermore, MDR is limited to a univariate model for interactions and family samples into a more powerful analysis. I propose here a comprehensive conceptual framework, referred to as generalized MDR (GMDR), for a systematic extension of MDR. The proposed approach is quite versatile, not only allowing for covariate adjustment, being suitable for the analysis of almost any trait type, e.g., binary, count, continuous, polytomous, ordinal, time-to-onset, multivariate, and others, as well as combinations of those, but also being applicable to various study designs including homogeneous and/or heterogeneous, case-only, and case-control studies. The proposed GMDR offers an important addition to the arsenal of analytical tools for identifying nonlinear multifactor interactions and unraveling the genetic architecture of complex biological traits. (This study is being supported by NIH grant DA025095.)

1844W

Meta-analysis identifies gene-by-environment interactions as demonstrated in a study of 4,965 mice. E. Kang1,2, H. Han1,3,4, N. Furuto1, J. Joo1,2, D. Shin1, R. Davis3, A. Lusis4, E. Eskin1,7, E. Eskin1,7. 1) Department of Computer Science, University of California, Los Angeles, CA, USA; 2) Division of Genetics, Brigham & Women’s Hospital, Harvard Medical School, Boston, MA, USA; 3) Partners HealthCare Center for Personalized Medicine, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Interdepartmental Program in Bioinformatics, University of California Los Angeles, CA, USA; 6) Department of Medicine, University of California, Los Angeles, CA, USA; 7) Department of Human Genetics, University of California, Los Angeles, CA, USA.

Identifying environmentally specific genetic effects is a key challenge in understanding the structure of complex traits. Model organisms play a crucial role in the identification of gene-environment interactions, as a result of the unique ability to observe genetically similar individuals across multiple distinct environments. A large number of model organism studies with varying environmental conditions measure the same traits. For example, knockout or diet-controlled studies are often used to examine plasma cholesterol levels in mice. These studies when examined in aggregate provide an opportunity to identify genomic loci bearing environment-dependent effects. However, straightforward application of traditional methodologies to aggregate data is often problematic. First, environmental conditions are often variable and do not fit the standard univariate model for interactions. Additionally, applying a multivariable model results in increased degrees of freedom and low statistical power. In this paper, we propose a meta-analytic approach based on a random effects model to identify loci involved in gene-environment interactions that pass a meta-effect threshold of p<0.001. The proposed approach is quite versatile, and can be used to study interactions in varying environmental conditions. Our approach is motivated by the observation that methods for discovering gene-environment interactions are closely related to random effect models for meta-analysis. We show that the interactions can be interpreted as heterogeneity and can be detected without utilizing the traditional uni- or multi-variate models for discovery of gene-environment interactions. We apply our method to combine 17 mouse studies containing a total of 4,965 animals. Using these studies, we identify 23 significant loci involved in HDL cholesterol levels. Among these, 23 loci are previously confirmed to have an effect on HDL cholesterol or closely related lipid levels in the blood, while 3 loci are novel. Several of these loci show significant evidence of gene-environment interactions in different environmental conditions. This approach focuses on a focus on a locus on chromosome 8 showing strong evidence of sex-by-environment interaction. Future work will consider interactions involving multiple environmental conditions.

1845T

GMDR: A conceptual framework for detection of multifactor interactions underlying complex traits. X. Lou. Dept Biostatistics, Univ Alabama Birmingham, Birmingham, AL.

Biological outcomes are governed by interacting networks consisting of multiple genetic and environmental factors that jointly act in often unpredictable ways. Determining multifactor interactions underlying complex traits is both the ultimate goal of human genetics studies but presents enormous statistical and mathematical challenges. The computationally efficient multifactor dimensionality reduction (MDR) approach, originally for a case-control study, has recently emerged as a promising tool for meeting these challenges. On the other hand, complex traits are expressed in various forms such as categorical, ordinal, continuous, count and time-to-event and have different data generation mechanisms that cannot be appropriately modeled by a dichotomous model; the subjects in a study may be recruited across a family, extending beyond its own analytical goals, research strategies and resources are therefore available, not only unrelated individuals from a homogeneous population. Although several modifications and extensions of MDR have in part addressed the practical problems, they are still limited in statistical analyses of diverse phenotypes, multivariate phenotypes and correlated observations. Furthermore, MDR is limited to a univariate model for interactions and family samples into a more powerful analysis. I propose here a comprehensive conceptual framework, referred to as generalized MDR (GMDR), for a systematic extension of MDR. The proposed approach is quite versatile, not only allowing for covariate adjustment, being suitable for the analysis of almost any trait type, e.g., binary, count, continuous, polytomous, ordinal, time-to-onset, multivariate and others, as well as combinations of those, but also being applicable to various study designs including homogeneous and/or heterogeneous, case-only, and case-control studies. The proposed GMDR offers an important addition to the arsenal of analytical tools for identifying nonlinear multifactor interactions and unraveling the genetic architecture of complex biological traits. (This study is being supported by NIH grant DA025095.)
1846F
Novel statistical framework for gene-environment interaction. A.R. Stef-
anescu, X. Wen. Biostatistics, University of Michigan, Ann Arbor, MI.
Investigations by Voorman, et al. (2011), Zuk, et al. (2011), and Wen and
Stephens (2011) have shown that model misspecification in the context of
a gene-environment interaction can lead to an array of serious analytical
issues, including missing heritability and inflated type I error. Firstly,
we demonstrate through simulation the detrimental effects of ignoring gene-
environment interaction in cases when interaction is known to be present.
Secondly, we address the weaknesses of the naïve model that ignores
interaction by comparing it to a traditional interaction model and also propo-
ing a more robust subgroup analysis framework. Finally, we intend to explore
the potential applications of the new method by integrating tissue-specific
eQTLs into GWAS analyses of complex traits.
Our preliminary results showed that when error variance was held constant
and genetic effect size was varied by environmental category in simulated
data sets, ignoring interaction explained little of the observed variance, while
the interaction models explained a much greater proportion of observed variance
as heterogeneity of genetic effect increased. Of the two interaction models,
the subgroup analysis performed the best, showing consistently higher
estimated heritability values than the traditional interaction model,
regardless of levels of heritability. Furthermore, when we allowed pheno-
type variance to differ by environmental category (a common phenomenon
observed at the cellular level), the interaction models also performed sub-
stantially better than the model ignoring interaction. Of the two interaction
models, the subgroup analysis performed the best. This result comes as
no surprise since the traditional interaction model assumes constant error
variance across environmental conditions.
The novel subgroup analysis structure is immediately applicable to studies
in which the environment variable is discrete, such as in different tissue
types (NIH GTEx Project, for example) and in threshold exposures. We
further intend to generalize this model to the continuous environment case.
The generalized model would be applicable to studies that use biomarkers
such as hormone levels as environmental variables.

1847W
Accounting for population structure in gene-by-environment interac-
tions in genome-wide association studies using mixed models. J. Sun,
W. Yang1, E. Kostem1, N. Furlotte1, D. He1, E. Eskin1,2. 1) Computer Sci,
Univ California, Los Angeles, Los Angeles, CA; 2) Department of Human
Genetics, University of California, Los Angeles, California, USA.
Genome-wide association studies (GWASs) attempt to identify genetic
variants associated with complex traits by collecting a set of unrelated individ-
uals. If individuals are related or from different populations, a phenomenon
called population structure may cause spurious associations. GWASs on
model organisms such as inbred mouse strains that are highly diverse in
genetic background are readily susceptible to population structure. Several
methods have been proposed to address this problem, and although GWASs
have identified numerous variants associated with many traits using the
methods, those variants explain only a small fraction of trait heritability.
Among several contributions that may account for a significant fraction of
heritability, one is gene-by-environment interactions (GEIs). Discovering
GEIs can provide insight into disease pathways, an understanding of the
effect of environmental factors in disease, pharmacokinetics and personal-
ized therapies. GEIs are often discovered in model organisms because of
possibility to manipulate different environments.
In this paper, we first show analytically that for the same reasons that
population structure causes spurious associations of genetic variants, it
causes spurious GEIs. We then use both simulation and inbred mouse
strains termed Hybrid Mouse Diversity Panel (HMDP) to observe spurious
GEIs. HMDP consists of 100 classical inbred and recombinant inbred (RI)
strains, and they are genotyped through the Mouse Genome Informatics
Project (MGIMP). We analyze their lipid phenotypes, and the environment is a thygocollate injection to recruit macrophages. We observe that test statistics for GEIs are inflated across most phenotypes, leading to spurious GEIs. Lastly, we propose a method to account for population structure for GEIs using mixed models. It has been shown that models designed to control population structure on effects of genetic variants. We show that directly applying mixed models
to GEIs, however, does not control population structure on GEIs because
current mixed models are designed to control inflation of genetic effects.
We extend our method to control population structure on both genetic
and GEIs effects and show that our method removes the inflation of test
statistics on simulation and the HDMP dataset.

1848T
Improved detection of variants with main or interaction effects using
a robust location-scale testing framework. D. Soave1,2, A.D. Paterson1,2,
L.J. Strug1, L. Sun1,2. 1) Division of Biostatistics, Dalla Lana School of
Public Health, University of Toronto, Toronto, Canada; 2) Research Institute,
The Hospital for Sick Children, Toronto, Canada; 3) Department of Statistical
Sciences, University of Toronto, Toronto, Canada.
The most common approach in genetic association studies is to test for
phenotypic mean (location) differences between genotypes. Complex
gene etiologies such as GxG and GxE interactions can result in phenotypic
variance (scale) differences between genotypes for a SNP of interest. For-
to date, no specific methods have been proposed to detect these interactions, but may not be practical due to issues such as multiple hypothesis testing or missing information concerning interacting exposures. As an alternative, Levene’s test for equality of variance has been proposed (Pare et al., 2010). Not surprisingly, when the distribution of a quantitative trait per genotype differs
mainly in variance, this scale-test has better power, but it is not robust for
detection of differences in location. More recently, Aschard et. al. (2013)
proposed a distributional test that compares the percentiles of phenotypic
values between genotypes. It has the advantage of detecting either mean or
variance differences, or both. While this method comprehensively evalu-
ates the phenotypic distribution between genotypes, the sample size
required to differentiate distributions is much larger than for detecting mean or
variance differences. Furthermore, the full information contained in a
(‘approximate’) normally distributed trait is well captured by the mean and
variance. We propose a joint location-scale testing framework. We consider
a direct likelihood approach that tests the null hypothesis of equal mean and
equal variance between genotypes. We compare findings with alterna-
tive Fisher’s and quantile-quantile plots of a set of traits observed in the
complementary individual location and scale tests. The combined methods
are more robust to model assumptions, and they allow for flexibility with the
specific individual tests chosen. Extensive simulation studies confirm that
the proposed location-scale testing framework is powerful for detection of
differences in location or scale both. Furthermore, in most of the simulation
scenarios considered by Aschard et. al. (2013), the proposed method is
equally or more powerful than the distributional test. Application to a candi-
date modifier study of cystic fibrosis and a genetic association study of type 1 diabetes complications show that the new testing framework
preserves the priority of previously identified top ranked variants while point-
ing to new candidates.

1849F
Improved detection of genetic exposures with unspecified effect modi-
fiers. T.L. Edwards1, C. Li2. 1) Department of Medicine, Vanderbilt Univer-
sity, Nashville, TN; 2) Department of Biostatistics, Vanderbilt University,
Nashville, TN.
Complex phenotypes often result from interplay of multiple genetic and
environmental factors. Association analyses can gain power by modeling
interaction effects when they exist. However, existing methods require
effects specification, effect modifiers, and exhaustive pairwise scans of
SNPs introduce well known computational, statistical, and logistical chal-
lenge. Knowledge of the loci that are contextually related to traits through
factors such as lifestyle, race, and drug treatment effects. For continuous
phenotypes, we propose a single-look association method that accounts
for interaction through combining both marginal mean and variance as linear
functions of genotype. We base this approach on the observation that when
controlling for effects such as genotype and environment, then the conditional variance for each genotype. We
show that applying separate location- and scale-tests on a single phenotype
of the same effect modifier varies by genotype, and that this influences the conditional variance for each genotype. We
dervide marginal mean and variance as functions of genotype and show that
different distributions are more robust to model assumptions, and they allow for flexibility with the
specific individual tests chosen. Extensive simulation studies confirm that
the proposed location-scale testing framework is powerful for detection of
differences in location or scale both. Furthermore, in most of the simulation
scenarios considered by Aschard et. al. (2013), the proposed method is
equally or more powerful than the distributional test. Application to a candi-
date modifier study of cystic fibrosis and a genetic association study of type 1 diabetes complications show that the new testing framework
preserves the priority of previously identified top ranked variants while point-
ing to new candidates.

Posters: Statistical Genetics and Genetic Epidemiology
1850W Genome Wide Interaction Study of Dengue Shock Syndrome. L. GRANGE1,2, J.F. BUREAUX, S. SAKUNTABHAIRI1,2. 1) Functional Genetics of Infectious Diseases, Institute Pasteur, Paris, FRANCE; 2) Institute Pasteur, Paris VII, FRANCE; 3) Department of Medicine, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

Although Genome-Wide Association Studies (GWAS) have been successful in identifying more than 500 single nucleotide polymorphisms (SNPs) associated with a broad range of diseases, these SNPs explain only a part of the genetic component of these diseases. Up to now, these GWAS have been limited to detecting one SNP at a time. Thus, only SNPs with marginal effect have been detected, while other alleles that act primarily through a complex mechanism involving interactions with other genetic variants and environmental factors have yet to be discovered. The identification of these interactions is a very challenging and powerful statistical method for conducting genome-wide interactions studies (GWIS) are needed. We compared power and false discovery rate among existing programs for detecting genetic interactions using simulation datasets with different genetic interaction models. Because of their respective low false discovery rate and high power, we chose to use Plink to perform exhaustive epistasis search on real data and Model-Based Multi-locus Dimensionality Reduction (MBMDR) to validate the best signals obtained. Using the previously reported GWAS dataset of dengue shock syndrome (Khor et al., Nature Genetics, 2011), we performed an exhaustive epistasis search, using Plink’s fast-epistasis mode, on 481,544 quality controlled SNPs and identified 2 couples of regions that contributed repeatedly to the best hits of this study. 3 SNPs located in the same region on chromosome 1 (between positions 77,814,710 and 78,041,795 - Reference Genome Build 36.3) gave strong interaction signals (P>10E-10). Although this region is approximately a unit interval graph, where the Max-Cut problem admits a linear-time solution. To phase the entire cohort, we iterate phasing of each G, updating the information regarding identity-by-descent segments after phasing so that \((u_i, v_j) \) is probabilistically assigned to the possible phases of \(G \). When sufficiently many segments are available, these iterations quickly converge to the correct phasing.


Long range haplotype phasing of heterozygous genotype calls had been proposed by considering segments of identity by descent to homozygous individuals. We hereby extend this framework theoretically by defining a special graph structure across a genomewide-typed or sequenced cohort. Formally, each segment of identity by descent is defined by a pair of typed individuals \(i,j \) and a genomic interval \((l,r)\). We define a graph \(G \) for each individual whose nodes are the segments \(\{[u_i, v_j(l,r)] \} \) and \(\{u_i(l,r) \} \) are linked by an edge if their corresponding segments overlap \(l<r \). Such edges are weighted according to the consistency of the footprints of these segments along the genomes of individuals \(i,j \) and \(k \) defined as log probability of their genotypes being identical by descent along the intersecting interval \((\max(l,v_j),\min(r,v_k))\). Maximum-likelihood phasing of \(i \) into her two parental haplotypes is equivalent to finding the maximum cut in \( G \). This framework allows handling alleles of various frequencies, incorporating sequencing/genotyping errors, while properly weighting interval length and missing data. Computationally, we observe that \( G \) is an interval graph, for which the Max-Cut problem is fixed-parameter-tractable, i.e. polynomial given a bound on the clique number of \( G \). Using a dynamic program that progresses along the maximal cliques in \( G \) along the genome. This implies a polynomial solution for maximum-likelihood phasing of \( i \). Practically, when \( G \) has large cliques, we observe that such cliques mean highly redundant phasing information, that we therefore omit. We further speedup performance by observing that most segments of identity by descent longer than a length threshold are only slightly longer than their defining cutoff. This means \( G \) is approximately a unit interval graph, where the Max-Cut problem admits a linear-time solution. To phase the entire cohort, we iterate phasing of each \( G \), updating the information regarding identity-by-descent segments after phasing so that \((u_i, v_j(l,r)] \) is probabilistically assigned to the possible phases of \(i \). When sufficiently many segments are available, these iterations quickly converge to the correct phasing.
1854T
Within-gene interactions in GWAS identifies novel susceptibility loci - The WTCCC data revisited. N. Sharif Eldin1, Q. Liu1, S. Jabban1, L. Wang2, C. Franco-Villalobos1, S. Mahasirimongkol1, H. Yansi1, L.J. Mar-tin1, K. Tokunaga3, Y. Yasui1. 1) School of Public Health, University of Alberta, Edmonton, Alberta, Canada; 2) College of Agriculture & Biotechnol-ogy, Zhejiang University, China; 3) Department of Human Genetics, School of International Health, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 4) Fukuiju Hospital, Japan Anti-Tuberculosis Association, Kiyose, Japan.

Genome-wide association studies (GWAS) examine single nucleotide polymorphisms (SNPs) associated with disease risk. A Standard single-SNP analysis, however, ignores combined effects of multiple SNPs; and SNP-set interactions remain largely unexplored at the genome-wide level. Here we show how exploring interactions of all SNPs within each gene can identify appreciable numbers of novel susceptibility loci. We re-analyzed six diseases in The Wellcome-Trust-Case-Control-Consortium (WTCCC) data: bipolar disorder (BD), coronary artery disease (CAD), hypertension (HT), rheumatoid arthritis (RA), type 2 diabetes (T2D), and type 1 diabetes (T1D). We considered two forms of SNP-set interactions: SNP intersection and SNP union. SNP-set interactions within each gene were assessed using logic regression. The number of genes showing strong evidence of associa-tion was: 88 for BD, 81 for CAD, 88 for HT, 161 for RA, 78 for T2D and 192 for T1D. All strong single-SNP signals of WTCCC and around 80% of recent GWAS meta-analyses signals were confirmed. In addition, strong evidence emerged implicating a large number of new discoveries supported by apparent biologically plausible links to disease. Top significant genes were: CBLN4 with BD, P2RX4 with CAD, BBOX1 with HT, STAG3L4 with RA, and RHOJ with T2D. Secondarily, 43 genes showed biologically plausible links to the six diseases in our interaction analysis. This emphasizes the importance of considering higher order SNP-set inter-actions in addition to the standard single-SNP analysis in GWAS.

1855F
Applications of Hidden Markov Models with Conditional Emission Probabilities to Identify Regions of Identity-By-Descent in Whole-Exome Sequencing Data. M. Kimmel1,2, S. Hicks3, S.E. Plon1. 1) Depart-ment of Statistics, Rice University, Houston, TX; 2) Department of Bioengin-eering, Rice University, Houston, TX; 3) Departments of Pediatrics and Human and Molecular Genetics, Baylor College of Medicine, Houston, TX.

Identifying regions of identity-by-descent (IBD) using the observed identity-by-state (IBS) status is an effective approach in identifying disease-causing mutations in Mendelian disorders. A previously developed inhomogenous first-order hidden Markov model (HMM) was applied to genotype data produced from whole-exome sequencing (WES) data to identify chromosomal regions of IBD in siblings with an autosomal recessive disorder [Bioinformat-ics 27: 829-836, 2011]. Our approach redefines the observed and hidden state space from a binary status to the true IBD and IBS status and uses inhomogeneous transition probabilities to account for position, distance and sex-specific recombination rates. To improve prediction accuracy, we extend the HMM to incorporate conditional emission probabilities and show these conditional emission probabilities vary as a function of the minor allele frequency. This analysis suggests minor allele frequency should be included in the determination of IBD regions when using WES data. We evaluate the HMMS using simulated human WES data and real datasets to identify regions of IBD. Using the known IBD status from the simulated families, we compare the root mean square error (RMSE) from each model averaged over a set of 100 simulated families. We show a first-order HMM with conditional emission probabilities using the redefined hidden IBD status has smaller RMSE = 0.17 compared to the first-order HMM previously developed (RMSE = 0.22). Our IBD model which incorporates minor allele frequency provides researchers a tool to filter large portions of the exome more accurately when searching for the causal variant(s) associated with Mendelian disease. Supported by CPRIT grant RP101089, NCI grant CA155767, NCI T32 training grant CA096520 and NCN (Poland) grant 519579938.

1856W
Cryptic relatedness in epidemiologic collections accessed for genetic association studies. J. Malinowski1, R. Goodloe1, K. Brown-Gentry1, D.C. Crawford2, S. Wang1. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Dept. of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN.

Over the last few years, epidemiologic collections have been a major resource for genotype-phenotype studies of complex disease given their large sample size, racial/ethnic diversity, and breadth and depth of pheno-types, traits and exposures. A major disadvantage of these collections is they often survey households and communities without collecting extensive pedigree data. Failure to account for substantial relatedness can lead to inflated estimates and spurious associations. To examine the extent of cryptic relatedness in an epidemiologic collection, we accessed the Third National Health and Nutrition Examination Survey (NHANES III), a popula-tion-based survey with DNA performed from 1991-1994 (n=7,159). Genome-wide genetic data is not available in NHANES III; though hundreds of SNPs genotyped in a variety of candidate genes are available for analysis. We performed identity-by-descent (IBD) estimates in three subpopulations of NHANES III: Hispanic (MEX, n=2073), non-Hispanic white (NHW, n=2631), and non-Hispanic black (NHB, n=2108) using PLINK to identify potential familial relationships from presumed unrelated subjects. After quality control, we calculated IBD with 784 SNPs in MEX, 721 SNPs in NHW, and 691 SNPs in NHB. In MEX, we identified two potential identical twin relationships (z-hat<0.90, p-hat<0.95), 312 potential parent/child relationships (z-hat<0.95, 0.48<z<0.55), and 98 potential full-sibling relationships (0.4<z<0.60, 0.15<z<0.35, 0.40<z<0.60). In NHW we did not observe any potential identical twin relationships; however, we did observe 1833 potential parent/ child relationships (0.20<z<0.35, 0.40<z<0.55), and 18 full-sibling relationships (0.40<z<0.60, 0.15<z<0.35, 0.40<z<0.60). We identified two potential identical twin relationships in NHB (z-hat<0.93, p-hat<0.95), 376 potential parent/child relationships (z-hat<0.85, 0.48<z<0.92), and 44 potential full-sibling relationships (0.40<z<0.60, 0.15<z<0.35, 0.40<z<0.60). We did not observe any 1st degree cousin relationships (0.70<z<0.80, 0.20<z<0.30) in any of the three subpopulations. We identi-fied numerous potential 2nd degree relationships (0.45<z<0.55, 0.23<z<0.33) as half-sibs, cousins, or nieces and nephews in each of the three subpopulations, though PLINK was unable to discrimi-nate among these relationships. Despite the lack of genome-wide data, our results suggest substantial cryptic relatedness in this epidemiologic collec-tion.

1857T
IBDLD-3: A fast and parallelized software package for IBD estimation from genomic data. L. Han, M. Abney. Human genetics, University of Chicago, Chicago, IL.

The estimation of identity by descent (IBD) from genomic data has gained much interest recently as a tool for a variety of genetic analyses. Here we present a new implementation of our IBD estimation method that is a dra-matic improvement in both speed and accuracy over our original software package. Computation time is substantially reduced due to, in part, the use of OpenMP to parallelize execution over many processors or cores. We have also implemented methodological improvements in our hidden Markov model based approach both for estimating the parameters of the model and in the computation of posterior IBD probabilities that result in both higher accuracy in the estimates and computational speed-ups. These improve-ments apply for both the case where pedigree information is available and when it is not. Our simulation studies show that IBDLD3 is much more accurate at estimating the true IBD sharing than the original implementation with reduced noise giving substantially improved performance at both point-wise estimates of IBD and in the detection of segments. IBDLD3 is robust to the genotype and pedigree error, or other forms of misspecified relationships, and is computationally fast enough to estimate all possible IBD sharing states at every SNP from a high-density genotyping array for hundreds of thousands of pairs of individuals. The open source C++ software package IBDLD v3.1 is freely available to be downloaded.
**1858F** Combinatorial Conflicting Homozygosity (CCH) enables the rapid identification of genetic linkage in the presence of multiple phenocopies. A.P. Levine1, T.M. Connor1, D.D. Ogur2, G.H. Neild1, P.H. Maxwell3, A.W. Segal1, D.P. Gale1. 1) Division of Medicine, University College London, London, United Kingdom; 2) Nicosia State Hospital, Burhan Nalbantoglu General Hospital, Nicosia, North Cyprus; 3) School of Clinical Medicine, Cambridge University, Cambridge, United Kingdom.

The analysis of multiplex kindreds has been successfully used to identify rare disease-linked variants; however genome-wide linkage analysis of such pedigrees can fail to identify linked loci in the presence of phenocopies, as can occur when the phenotype or index case is uninformative. To overcome these challenges, we developed a method based on the principle of conflicting homozygosity (CH) in which runs above a certain length of consecutive biallelic markers containing no occurrences of homozygosity for both major and minor alleles across a set of affected individuals are inferred to result from identical by descent (IBD) inheritance. An equivalent approach was described by Leibon et al. and Thomas et al. (2008), referred to as Shared Genomic Segment Analysis. We add a combinatorial feature to our method, packaged in a Python software program, in which the analysis is automatically repeated on all subsets of affected individuals. To characterise the parameter thresholds that define a locus as one likely inherited IBD, we analysed subsets of unrelated HapMap individuals. Gene-dropping simulations in pedigrees defined the sensitivity and specificity of our approach. We applied our program to a large family in which microscopic haematuria, proteinuria or renal impairment segregated in affected individuals. Gene-dropping simulations in pedigrees defined the sensitivity across a set of affected individuals and no loci shared IBD. We hypothesised the presence of one or more loci with suggestive or significant linkage. Analysis of all 12,376 combinations of 11 of the 17 affecteds (which took ~7 hours running serially) identified three loci, one of which contained the gene COL4A3, mutations in which are known to cause familial kidney disease. Sequencing identified the previously described pathogenic G871C mutation in all 11 affecteds carrying the haplotype, no unaffected individuals and none of the six affecteds who did not share the linked haplotype (including the proband). We propose that combinatorial CH can be used to identify loci shared IBD in the presence of phenocopies and may be of value in future genotyping.

**1859W** Investigating the Importance of Disparate Genetic Influences across African and European Descent Populations. J. de Candia1,2, M.C. Keller1,2. 1) Psychology & Neuroscience Dept, University of Colorado, Boulder, CO; 2) Institute for Behavioral Genetics, University of Colorado, Boulder, CO.

The degree to which genetic influences overlap across ethnicity has been a matter of contention. Besides the possibility that causal variants (CVs) differ between ethnicities, traits may be differentially predicted by genes due to different SNP-CV linkage disequilibrium (LD) patterns as well as artifactual differences due to age of onset or disease progression. We hypothesized that the factors that can affect genetic correlations tagged by SNPs between ethnic populations, we simulated CVs in 10,849 African descent (AD) and 10,007 European descent (ED) genotypes from four datasets (ARIC, WHI, MESA, CARDIA). A random set of 2,000 SNPs that passed quality control procedures and that were largely independent of each other were chosen as CVs for each ethnicity. For each CV, we constructed a continuous phenotype with a heritability of 0.5, and used a bivariate mixed linear model (Lee et al., 2012) to estimate the heritabilities and correlation tagged by nearby SNPs. We then explored the extent to which LD and CV allele frequencies, as well as differences in these parameters across ethnicity, predicted these parameters. Across CVs, heritabilities were underestimated for both ADs (0.40, sd=0.11) and EDs (0.44, sd=0.12), and as expected, this appeared to be be explained by incomplete LD (r2=33 in ADs and 17 in EDs, p < 1.55e-06). We found that the mean genetic correlation was 0.94, only slightly lower than its expected value of 1. Controlling for minor allele frequencies differences, differences in LD patterns across ethnicities lowered genetic correlations (r2 = 0.25, p < 2e-16). On the other hand, controlling for LD pattern differences, allele frequency differences did not predict differences in genetic correlations (r2 = 0.01, p > 0.05). Our findings suggest that genetic correlations will be high between ethnic groups when CVs are the same, despite differences in ethnic allele frequencies, SNP LD patterns, and MAFs. Thus, low genetic correlations between ethnic groups might suggest different causal variants are influencing the phenotypes and/or technical aspects of the design artificially reduce genetic correlations. Of the two factors investigated here, differences in LD patterns across ethnicities may have a greater influence on SNP-correlations than did differences in MAFs. We discuss the implications of these findings on genetic correlations we observe for height and BMI.


Genetic variants genotyped by commercial genotyping platforms capture approximately half of the estimated heritability of the dozens of complex human phenotypes examined to date (Visscher et al., 2012). Rare variants may make up a substantial fraction of the missing heritability (Manolio et al. 2009). Recent advances in sequencing technologies have enabled deeper exploration of rare genetic variants and their role in disease susceptibility than was previously possible with array based genotyping (Eichler et al. 2010). Although sequencing costs are likely to continue to fall, the resources to achieve the sample sizes necessary to perform a well-powered rare variant analyses can be prohibitive (Liu et al. 2013). In this work, we examine the use of segments of pairwise identity by descent (IBD) computed from array based genotype data to serve as a proxy for rare variants, preventing the need for expensive sequencing and permitting rare variant analyses of existing data sets. Our assumption is that the majority of rare variants have occurred recently and therefore exist on the same IBD background. We apply existing methods (B. L. Browning and Browning 2011) (Gusev et al. 2009) to identify segments of IBD and a clustering approach to group individuals with pairwise IBD status commensurate with an untyped rare variant (Gusev et al., 2011). Instead of testing clusters for association directly, we turn each cluster into a ‘pseudo-SNP’ which represents an untyped rare variant. Since traditional single-variant tests tend to be inadequately powered to detect associations with rare variants, various gene-based tests have been proposed to examine the combined effects of rare variants. We test genes for association using SKAT in combination with IBD-based pseudo-SNPs (Wu et al. 2011). We applied our method to two Latino cohorts, one consisting of asthmatics from the GALA cohort (Burchard et al. 2004) (Kumar et al. 2013) and the other consisting of breast cancer cases (Fejerman et al. 2008). We examined in detail the underlying assumptions of our method by determining the fraction of rare variants captured as pseudo-SNPs in several exome and whole genome sequencing data sets.

**1861F** Computationally-efficient long-range phasing with very large datasets. M.J. Barber1, R.E. Curtis2, K. Noto3, Y. Wang1, J.M. Granka1, N.M. Myres2, J.K. Byrnes1, C.A. Bahl1, K.G. Chahine1. 1) Ancestry.com, San Francisco, CA; 2) Ancestry.com, Provo, UT.

While computationally intensive, phasing a large set of genotypes (i.e. > 100,000 samples) into probable haplotypes presents the opportunity to leverage sample size to increase phased accuracy for every sample. Phasing switch error rate can be greatly minimized when using genotypes with a specified parent-offspring relationship, but such datasets are not universally available. Recent advances in methodology have utilized the large number of identity-by-descent segments (IBD-SEGS) between a sample and the rest of the samples in the dataset to improve phased accuracy. These ‘long-ranged’ phasing approaches use an IBD-SEG to help phase the matched IBD region. At AncestryDNA, we are applying the principle of long-range phasing to very large (and growing) datasets of genotyping data. Our approach assembles high-confidence IBD-SEGS to form an explicit surrogate parent for each sample. The assembly of an explicit surrogate parent is computationally efficient: the only requirement is assessing each IBD-SEG for quality separately, rather than performing a joint analysis. Given an explicit surrogate parent, phasing and IBD-SEGS can then be updated. Our approach has the added advantage of enabling updates to the phase estimates of full or partial genotypes when new high-quality IBD-SEGS are identified, as is common in constantly growing databases such as AncestryDNA's. We test the accuracy of our approach using simulated genotyping datasets and thousands of confirmed parent-offspring relationships from the AncestryDNA database. Our novel approach aims to efficiently and accurately phase large numbers of samples in a way that could be relevant and widely practical for a variety of applications, including linkage disequilibrium and association studies that are being generated by the genetics community.
1862W


Identity by descent (IBD) between two individuals means that their alleles are identical because they were inherited from a common ancestor. This information can be used via IBD mapping to increase the power of association analysis by grouping single nucleotide variants (SNVs) based on IBD. The basic principle of IBD mapping is to look for segments of DNA that are shared identical by descent more often among cases than controls.

HapFABIA is a biclustering algorithm that was originally designed to extract short IBD segments that are present in multiple individuals from large sequencing data using only rare variants. In this setting we used HapFABIA on SNP microarray data from the Autism Genetics Resource Exchange (AGRE) to look for IBD segments that are shared more often by cases than controls.

We found several IBD segments that were almost exclusively shared by cases. Some of these map to genes that have been previously associated with autism. Further analyses are needed to confirm or disprove these results.

1863T


The methods of principal component analysis (PCA) and hierarchical clustering on individual genotypic data have been widely used to detect population substructure in genome-wide association studies. The principal component axes often represent perpendicular gradients in geographic space, and we provide an interpretation of PCA based on relatedness measures, which are described by the probability that sets of genes are identical-by-descent (IBD). An approximately linear transformation between the projection of individuals onto the principal components and allele admixture fractions assuming two or more ancestral populations is revealed. Furthermore, a measure of individual dissimilarity based on the coancestry coefficient was proposed for hierarchical clustering. Compared to other dissimilarity measures (such as identical-by-state methods), its expected value is directly related to the kinship coefficient without being confounded by allele frequency, and it is a moment estimator and suited for large-scale GWAS data. Both PCA and hierarchical clustering were applied to HapMap Phase II and III data. The population admixture proportions inferred by PCA are consistent with what HAPMIX and ADMIXTURE estimate. Hierarchical clustering successfully separates Chinese and Japanese samples in HapMap Phase II data based on the coancestry coefficients. Finally, a combination of PCA and hierarchical cluster analysis should help us better understand population structure for isolated and admixed populations.

1864F


A large number of genetic studies involve samples from structured populations, including those with ancestry admixture. Estimating measures of relatedness such as kinship coefficients and identity by descent (i.b.d.) sharing probabilities in these samples is a challenge. Commonly used relatedness estimators, such as those implemented in PLINK, the genetic relationship matrix (GRM), and the EM algorithm, assume population homogeneity and are thus prone to mis-specification of population structure. The REAP method (Thornton et al., 2012) addresses relatedness estimation in admixed populations by utilizing individual specific allele frequencies. In order to estimate individual specific allele frequencies, however, REAP relies on estimates of admixture proportions and ancestral allele frequencies from likelihood based methods such as ADMIXTURE (Alexander et al., 2009) or FRAPPE (Tang et al., 2005). As a consequence, REAP is inherently sensitive to mis-specification of model assumptions in these methods; mis-specification of the number of ancestral populations represented in the sample, or insufficient reference samples from them, may lead to inaccurate admixture proportions and/or biased allele frequency estimates. We propose a model-free method of estimating i.b.d. sharing probabilities using principal components without the requirement of external reference samples. We accomplish this through the use of a modified GRM, similar to that in REAP, along with an analogous novel estimator for the probability of sharing two alleles i.b.d. that relies on an alternative coding of genotype values. We demonstrate the utility of our method by estimating i.b.d. sharing probabilities of more than 12 000 African Americans and Hispanics from the Women’s Health Initiative study.

1865W

Homozygosity mapping combined with linkage analysis in human families in the age of high-density DNA variants. J. Ott1,2, Y. Li3, E. Engle4, S. Shaaban3. 1 Institute of Psychology CAS, Beijing, Beijing, China; 2 Ott Lab, Rockefeller University, New York; 3 School of Statistics, Shanxi University of Finance & Economics, Taiyuan, Shanxi, China; 4 Departments of Neurology, Medicine, and Ophthalmology, Harvard Medical School, Howard Hughes Medical Institute, Boston, MA; 5 Department of Neurology, Harvard Medical School, Dubai Harvard Foundation for Medical Research, Boston, MA.

Exome sequence data are often obtained in small human families. Extracting SNPs from variant (vcf) files allows for linkage analysis but power is often low. Also, multipoint linkage analysis tends to be problematic due to correlations among closely spaced SNPs. Here we propose a modern, new approach: (1) At each SNP, only one lod score (at t = 0.0001) is calculated. (2) With SNPs being in chromosomal order, a run of positive lod scores (ROL) is defined as a set of contiguous SNPs all with lod scores > 0.01, delimited by SNPs with lod scores ≤ 0.01 or a chromosome end. ROLs are analogous to runs of homozygosity (ROH) in homozygosity mapping in unrelated individuals but are more realistic as they take into account allele frequencies and family relationships. (3) For two families with a given disease, we focus on long ROLs (eg. > 500 bp) and count the number N of ROLs between the two families that overlap by k base pairs, eg. k = 1000 bp. (4) We approximate the null distribution of N by computer simulation by randomly placing observed ROLs on chromosomes so as to obtain a p-value associated with N. We applied our new method to two 2-generation families with comitant strabismus, each with 1 or 2 affected children and two related parents. Traditional linkage analysis furnished maximal lod scores < 2 (genome-wide p > 0.05). Focusing on the 80 longest ROLs in each family, our approach identified N = 17 ROLs overlapping by at least 1 bp (p = 0.1319), and N = 16 ROLs overlapping by at least 1 Mb (p = 0.0027). Thus, this approach finds strong genome-wide significance where traditional linkage analysis is unsuccessful, and at the same time allows replicating linkage results in a second family, with both families being of only modest size. As negative controls, (1) we applied our method to two families with very different diseases, with resulting p-values > 0.50; (2) also, when our ROLs are all singletons, our approach performed exactly as expected. ROLs significantly overlapping between two families presumably harbor common susceptibility variants. Thus, by design, this procedure takes into account multiple disease genes, possibly located on different chromosomal regions, and our new method is expected to be more powerful for recessive than dominant acting variants but this is not a serious shortcoming as most mutations are recessive.

1866T


Variant detection from genome-wide sequencing data is essential for the analysis of disease causing mutations and elucidation of disease mechanisms. However, variant calling in low coverage regions is difficult due to sequence read errors and mapping errors. Hence, variant calling approaches that are robust to low coverage data are demanded. We propose a new variant calling approach that considers pedigree information and haplotyping based on sequence reads spanning two or more heterozygous positions termed phase informative reads. In our approach, genotyping and haplotyping by the assignment of each read to a haplotype based on phase informative reads are simultaneously performed. Therefore, positions with low evidence for heterozygosity are rescued by phase informative reads, and such regions with low evidence for haplotyping in a synergistic way. In addition, pedigree information supports more accurate haplotyping as well as genotyping, especially in low coverage regions. Since undesirable influence from homozygous positions to the read assignment prevents accurate haplotyping, we introduce latent variables that determine zygosity at each position, and avoid the influence from homozygous positions by using the latent variables. In performance evaluation with a parent-offspring trio sequencing data, our approach outperformed existing approaches in accuracy on the agreement with SNP array genotyping results. Also, performance analysis considering distance between variants showed that the use of phase informative reads is effective for the accurate variant estimation, and further performance improvement is expected with longer sequencing data.
Gender differences in intelligence by mathematical model explains gender differences in intelligence by variation in mating preferences. M. Nagel, Molec Gen, Ctr Nephrology, Weissswasser, Germany.


The application of pathway and gene-set based analyses to high throughput data is increasingly common and represents an effort to understand underlying biology where single-gene or single-marker analyses have failed. Many such analyses rely on the a priori identification of genes associated with the trait of interest. In contrast, this variance-component based approach to the analysis of microarray data identifies individual differences in the variance of expression of genes within each pathway. Previously generated mRNA data from 1240 Mexican-American participants in the San Antonio Family Heart Study was considered in relation to circulating HDL levels. Previous single-gene and over-representation analyses of HDL in this sample identified one gene and four associated pathways. Transcript levels were ascertained from total RNA isolated from lymphocytes and analyzed with the Illumina Sentrix Human Whole Genome (WG-6) Series I BeadChips. All probes were standardized by z-scoring within individuals. 16,681 probes were detectable above baseline levels in the majority of individuals, have non-zero heritability, and can be annotated with ReMOAT mapping. Expression levels for genes composing 228 pathways drawn from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database were adjusted for age and sex before inverse normalizing. For each pathway, the normalized values were used to construct matrices of Malhanobis distances between each pair of individuals in R. Using SOLAR, a polygenic null model of HDL level was generated including sex and age as covariates. A model including each distance matrix in turn was compared to the null model to determine if the pathway is significantly associated with HDL. A final model was constructed by iteratively including each of the associated pathway matrices, in order of magnitude of effect, and removing matrices that do not explain additional variation, likely due to overlap in genes between the pathways. This method has a number of advantages when compared to single transcript and pathway over-representation analyses, including the ability to estimate the proportion of variation explained by each pathway and all pathways combined, as well as the logistical advantage of only calculating the distance matrices once for each mRNA data set regardless of the number of phenotypes. Most importantly, this method allows for the simultaneous consideration of multiple pathways to determine if they represent independent associations.

1867F Genetic pathways enriched with type 1 diabetes suggest novel causal genes for type 1 diabetes. M. Evangelou, D.J. Smith, O.S. Burren, N.M. Walker, J.A. Todd, C. Wallace. JDRF/WT Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge, Cambridgeshire, United Kingdom.

Pathway analysis can complement single-SNP analysis in exploring genome-wide association (GWA) data to identify disease associated genes. We applied a gene-based competitive pathway analysis to previously published type 1 diabetes (T1D) GWA data covering a total of 5,916 T1D cases and 7,338 controls, all of white European origin. After mapping each SNP to a unique gene according to distance, we explored three different statistics that summarize the association of each gene (the minimum single SNP p-value, the mean -log p value, and the Fisher’s product statistic) and two methods for summarizing the association across genes in a pathway (Fisher’s product method and the adaptive rank truncated product method). Of 314 BioCarta and 1,272 Reactome pathways, 30 and 101, respectively, were associated using at least one statistic, and just five (three BioCarta and two Reactome) were associated using all six statistics. The BioCarta pathways included L-2 receptor beta chain in T-cell activation, ‘Antigen dependent B-cell activation’ and ‘The Co-stimulatory signal during T-cell activation’ and the Reactome pathways were ‘GRB7 events in ERBB2 signaling’ and ‘Regulation of IFNG signaling’. This echoes much that is known about the etiology of type 1 diabetes.

We identified SNPs with small GWAS p-values near genes in associated pathways that have not been reported as associated with T1D. The genes FASLG, RAF1 and SOCS1 were identified as potential novel gene risks of T1D. The FASLG is associated with Celiac disease but has not reported as associated with T1D, whereas the SOCS1 gene and with a number of autoimmune diseases including T1D. The RAF1 gene has not been reported as associated with T1D or with any other autoimmune disease. SNPs near these three genes are currently being genotyped in a replication cohort. Such replication will prove the utility of pathway analysis not only for identifying pathways of interest, but for using this information to alter our prior belief of association and thus identify novel disease associated SNPs that do not reach stringent genomewide significance levels.


Prediction of the consequences of intervention effects, such as gene knockdowns, from observational data is of central interest in biomedical sciences. Experimental perturbations are expensive and sometimes impossible due to ethical consideration. This motivates the development of statistical and causal methods to prioritize between experiments. Graphical models together with intervention calculus provide a formal framework for predicting bounds on intervention effects from observational data but only when the underlying graphical model is known. Recently developed methods have shown how the graph itself can be inferred from data, with the limitation that only bounds on the causal effects can be estimated (Maathuis, M. H. et al., Estimating high-dimensional intervention effects from observational data. Annals of statistics, 2009). These methods have been shown to outperform (non-causal) statistical methods in terms of predicting intervention effects, and also been validated through perturbation experiments. These methods are, however, based on accurate inference of the equivalence class of a Directed Acyclical Graph (DAG) from data. Here we present an extension to previously reported methods that allow for uncertainty in the inferred DAG, combining ideas from the area of causal inference with a Bayesian treatment of the structure of the graphical model. Our objective is to properly account for the uncertainty associated with the structure of the graphical model that is inferred from molecular profiling data. The proposed method utilises Markov Chain Monte Carlo sampling to order combinations of models using the transmission algorithm and averages over models, with Bayesian model averaging to approximate the posterior of intervention effects. We report simulation results that demonstrate under realistic conditions the proposed methods have considerable advantages over existing methods, as well as results from applications to real omics data in the context of cardiovascular disease.
1871W
A Powerful Statistical Method for Genetic Pathway Analysis. N. Liu, Q. Yan, N. Yi. Dept Biostatistics, Univ Alabama at Birmingham, Birmingham, AL.

Genetic pathway represents certain biological mechanisms of diseases. Pathway analysis takes into account the biologically pathway-gene-marker hierarchical structure, while traditional analysis strategy for genome-wide association studies (GWAS) usually focuses on single marker analysis. Therefore pathway analysis may potentially complement single-marker analysis and provide additional insights for the genetic architecture of complex diseases, thus may lead to higher statistical power and more biologically meaningful interpretation. In this work, we proposed a new method for genetic pathway analysis. It is based on kernel machine testing and adaptive rank truncated product test. The genotype data of 2000 subjects from WTCCC Type 1 Diabetes were used for simulation studies. We compared the performance of our new method and several other methods. The preliminary results show that the new method has the highest power while maintaining correct Type I error rate in most of the scenarios we considered.

1872T
A supervised dimension reduction approach for pathway-based analysis in Genome-wide association study. Z. Wei1, J. Li², W. Wang1, H. Hakonarson1. 1) New Jersey Institute of Technology, Newark, NJ; 2) Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, PA.

Pathway-based gene set enrichment analysis (GSEA) has been routinely conducted as a complementary approach for conventional single-SNP based association tests in genome wide association study (GWAS). Since the first simple extension from the GSEA for microarray gene expression data years ago, methodology research is undergoing rapid development for improving GSEA in GWAS. Quite a few new methods have been proposed. One common strategy is to utilize dimension reduction technique, hoping to improve power by reducing the unnecessary large degree of freedom. However, most of existing methods are unsupervised approaches, namely, not exploiting trait information, when reducing dimensionality. Here we employ a supervised strategy by applying the ridge regularized Kernel Sliced Inverse Regression (KSIR) to achieve dimension reduction. Not only can KSIR exploit trait information, but also it is capable of capturing non-linear effects by flexibly employing various non-linear kernel functions. Using simulation studies, we show that the KSIR method outperforms conventional unsupervised competing methods in terms of causal pathway ranking and the statistical power. We also demonstrate the superior performance of KSIR in analysis of a real dataset, the WTCCC Ulcerative Colitis dataset consisting of 1782 cases and 3773 controls as the discovery cohort, and 591 cases and 1639 controls as the replication cohort. We identify several immune and non-immune pathways relevant to Ulcerative Colitis, including some novel ones that may be worthy of further investigation for their roles in the development of Ulcerative Colitis.

1873F
Rank-based analysis of transcriptome data reveals biologically relevant atopic dermatitis genes and pathways. TB. Marsha1, D. Ghosh2, GK. Hershey1, JA. Bernaile2. 1) Division of Asthma Research, Cincinnati Children’s Hospital Medical Center, University of Cincinnati, Cincinnati, OH; 2) Immunology and Allergy, Department of Internal Medicine, University of Cincinnati, Cincinnati, OH.

Atopic Dermatitis (AD), a chronically relapsing inflammatory disorder of the skin, affects an estimated 15 to 30% of children and 2 to 10% of adults. While the exact causes of AD are not fully understood, genetic differences may also be responsible for AD susceptibility. Several independent groups have identified differentially expressed genes (DEGs) across versus AD patients using microarray technology. However, there is little overlap in the DEGs lists reported by different groups due to variation from random noise, biological and experimental differences, as well as differences in the extraction and handling of RNA samples. Therefore, a statistical rank-based analysis is necessary to identify a set of genes that are consistently dysregulated among multiple independent microarray studies. We obtained AD-related expression data from the GEO dataset (http://www.ncbi.nlm.nih.gov/geo/; accessed in March 2013) using ‘human’ (organism) AND ‘Atopic dermatitis’ search terms. In this study we focused on six datasets after eliminating data obtained from non-human/ex-vivo cell-culture experiments. We, then, screened genes differentially expressed between skin samples from AD and healthy control subjects using classical t-test and p-values of <0.05 as significant threshold. We identified differentially expressed genes (with fold change >1.5) consistently up/down-regulated in at least 5 out of 6 datasets and ranked them according to their average fold changes. The associations of particular gene across studies may represent a true association with AD and could be given highest priority. We identified 144 DEGs, of which 8 were found to be consistently upregulated, while 16 genes were consistently downregulated in all the datasets analyzed. Top upregulated genes were KRT16 (related to Keratinocyte production) and LTF (related to the expression of lactoferrin, as anti-microbial protein), while the top down-regulated genes were LOR (loricrin), LCE2B (late cornified envelope 2B), FLG (filaggrin) genes- mostly related to skin barrier function. Pathway analysis revealed that the DEGs are related to epidermal development or inflamma- tory pathways. Thus, rank-based analysis of publicly available data indicates biologically relevant genes and pathways thus providing mechanistic clues to observed gene expression patterns in AD.

1874W
Novel Top Down Dissection of Complex Traits Based on Germline Signatures. H.K. Im1, N.J. Cox2. 1) Department of Health Studies, University of Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL.

Genome-wide association studies and more recently whole genome/exome sequencing studies have identified thousands of genetic variants reproducibly associated with many complex diseases and traits. However, by and large the biological mechanisms underlying these discoveries are still lacking. In addition, it is becoming increasingly clear that a large number of variants with modest effect sizes are contributing a substantial proportion of the heritability of complex traits. Thus methods that aggregate variants into biologically meaningful units and integrate functional and regulatory evidence are dearly needed. Several gene-based approaches have emerged in order to address this problem. We propose to go a step further with a top down systems level approach that dissects the anatomy of complex traits in terms of higher level biological processes and components. For this purpose, we are in the process of developing a catalog of germline signatures (GSigDB) using cellular, animal, and bioinformatic models as well as clinical and non-clinical human data. These signatures allow testing novel hypotheses regarding disease etiology, pleiotropy among traits, to name a few. These signatures are stored as optimal weights to be applied to whole genome data and generate polygenic scores or predictions of endophenotypes. The weights are computed integrating data from existing large scale meta analysis and regulatory and functional information. For example, for the insulin resistance signature we use the effect sizes from the meta analysis of ~50K individuals published by the MAGIC consortia. We applied our approach to the 7 diseases from Wellcome Trust Consortium. We found that both insulin resistance (p=0.0029) and beta cell function (p=0.053) have similar effect sizes on the risk to type 2 diabetes. Interestingly, we also found that type 1 diabetes risk shows a significant association with insulin resistance (p=0.054) and beta cell function (p=0.040) with similar effect sizes. BMI was not significantly associated with type I diabetes, as expected. For the 5 other diseases there was no significant association with insulin resistance or beta cell function except for a negative except for a negative association with Crohn’s disease and insulin resistance (p=0.0057). This finding is intriguing but needs to be replicated in an independent dataset. In conclusion, our top down approach allows us to ask high level biological questions and generate new mechanistic hypothesis.

Posters: Statistical Genetics and Genetic Epidemiology 187
1875T
A systems-biology approach to identify and prioritize sub-networks of functionally-related genes for Alzheimer’s disease and subsequent in vivo validation of candidate genes using a C. elegans model of Aβ toxicity. S. Mukherjee1, M. Kaeberlein2, J. Kauwe3, A. Najj4, P. Crane4, Alzheimer’s Disease Genetics Consortium. 1) Department of Medicine, University of Washington, Seattle, WA; 2) Department of Pathology, University of Washington, Seattle, WA; 3) Departments of Biology and Neuroscience, Brigham Young University; 4) Perelman School of Medicine, University of Pennsylvania.

Background: Recent genome-wide association studies (GWAS) have identified around 20 variants as late-onset Alzheimer’s disease (LOAD) susceptibility loci in whites. In addition to these single loci tests, it is important to detect and understand combined effects of multiple associated genes on LOAD. We performed a preliminary network analysis incorporating human protein-protein interaction database from BioGRID to the HapMap2-imputed combined ADGC data set. Post-GWAS, this helps researchers to prioritize functionally related genes and networks that are of the highest biological relevance underlying the pathogenesis of LOAD. Methods: We combined HapMap2-imputed data sets from 15 studies after performing strict quality control. We performed a case-control association for LOAD adjusting for HapMap2-imputed data sets from 15 studies after performing strict quality relevance underlying the pathogenesis of LOAD. Methods: We combined ADGC data set. Post-GWAS, this helps researchers to prioritize candidate genes or sub-networks for LOAD. We then attempted to functionally validate candidate genes from this network in vivo using a transgenic C. elegans model of Aβ42 toxicity. Network analysis and validation in the C. elegans model of Aβ42 toxicity. Conclusions: We were able to identify a set of significant modules and candidate genes, including some well-studied genes not detected in the single-marker analysis of GWA studies for LOAD, and to demonstrate a role for two of these genes as modifiers of Aβ42 toxicity. This in vivo work supports a complementary data to a GWAS of a complex disease phenotype by incorporating biological knowledge derived from protein-protein interactions and allows for initial functional validation in vivo. Further functional enrichment analysis will define whether these novel loci may provide targets for interventions to ameliorate LOAD.

1876F

Understanding the regulatory interactions between genes, and identifying those interactions that are shared across human tissues and those that are specific to each tissue, is essential for understanding regulatory mechanisms that govern similarities and differences between tissues and can help reveal tissue-specific responses to environmental or genetic perturbations. The GTEx project, with RNASEq data from more than 1,000 samples across 35 human tissues, provides an opportunity for such an analysis. Here we utilize the GTEx dataset to learn regulatory networks across these 35 tissues, where each network captures interactions between genes in a particular tissue. To do so, we present a method that makes use of a tissue hierarchy, enabling us to learn such networks even for tissues with very few samples. Our method consists of two components. First, we derive a tissue hierarchy that reflects the relationships between tissues based on expression data using hierarchical clustering, and is shown to be consistent with tissue lineage relationships. We then develop a fast algorithm (1) that learns shared regulatory networks in a hierarchical manner to learn sparse regulatory networks in each tissue, using a L2-regularization penalty to enforce the fact that tissues nearby in the hierarchy should have similar networks. This regularization increases accuracy by allowing for transfer learning from tissues for which we have many samples to tissues for which we have few. We show, using cross validation on three different gene sets, that this algorithm learns networks more accurately than either learning a single network for all tissues or learning a network for each tissue independently. We discuss the broad applicability of this algorithm to any set of hierarchically related networks, such as those from phylogenetic or cancer datasets. We further validate the accuracy of our learned networks by comparison to the GTEx project (including the topologically similar datasets). This analysis is needed to determine whether these novel loci may provide targets for interventions to ameliorate LOAD.

1877W
Seasonal changes in gene expression represent cell type composition in whole blood. M. Neeleman1, S. de Jong1, J.J. Luykx2, M.J. ten Berg3, E. Strømgren4, H.H. ten Breeuwer5, L.C. Stijvers6, J.E. Buiter-Voskamp7, S.C. Bakker8, R.S. Kahl9, S. Horvath10, W.W. van Solinge11, R.A. Ophof9,11,12,13. 1) Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands; 2) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA; 3) Department of Psychiatry, ZNA hospitals, Antwerp, Belgium; 4) Department of Clinical Chemistry & Haematology, Division of Laboratories & Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 6) Department of Pharmacoeconomics and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences; Faculty of Science, Utrecht University, Utrecht, The Netherlands; 7) Department of Clinical Pharmacy, University Medical Center Utrecht, The Netherlands; 8) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA; 9) Department of Biostatistics, School of Public Health, University of California, Los Angeles, California 90095, USA.

Seasonal patterns in behavior and biological parameters are widespread. Here, we examined seasonal changes in whole blood gene expression profiles of 233 healthy subjects. Using weighted gene co-expression network analysis, we identified three co-expression modules showing circannual patterns. Enrichment analysis suggested that this signal stems primarily from red blood cells and platelets. Indeed, a large clinical database with 51,142 observations of blood cell counts over three years confirmed a corresponding seasonal pattern of counts of red blood cells, reticulocytes and neutrophils. We further showed that seasonal changes in gene expression profiles in whole blood represent biological and clinical relevant phenomena. Moreover, our findings highlight possible confounding factors relevant to the study of gene expression profiles in subjects collected at geographical locations with disparaging seasonality patterns.

1878T
Imputation Performance of ~4,000 genomes from the UK10K Project. J. Huang1, B. Howie2, J. Marchini3 on behalf of the UK10K Cohorts Consortium. 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Department of Human Genetics, University of Chicago, Chicago, Illinois, USA; 3) Department of Statistics, University of Oxford, Oxford, UK.

Imputation from reference haplotypes has proved instrumental to the success of genome-wide association studies. Imputation based on whole-genome sequencing (WGS) data is expected to further increase the utility of this approach to variants of lower minor allele frequency (MAF). To date, the 1000 Genomes project has made publically available data for a set of 1,092 genomes (4X) for individuals of multiple ethnicities. The UK10K Cohorts project has released a dataset of 3,781 genomes (6X), aiming to exhaustively characterize genetic variation down to 0.1% MAF in the British population. We compared the UK10K reference panel with the 1000 genomes panel alone and the two combined. The combination of two WGS reference panels was carried out using a new functionality in IMPUTE2. We generated two pseudo-Illumina 610k bead chip datasets as imputation targets by subsampling WGS data from: (1) 10% samples of the UK10K project; (2) four European samples from Complete Genomics (80X). The imputation quality was measured by Pearson correlation coefficient (r2) between the imputed genotype dosage and the masked sequences. For each of the above six scenarios (three reference panels, two SNP-arrays), we run two further analyses: (1) comparing imputed WGS data using a new algorithm implemented in SHAPEIT2; (2) varying the -k hap parameter in IMPUTE2 that controls the number of haplotypes used to impute each sample (500 vs. all haplotypes).

Comparison of 3,781 UK10K genomes significantly improved imputation quality for variants with MAF between 0.1% and 5%. Increasing the number of haplotypes sampled and re-phasing the WGS reference panel provided further improvement. Adding 1000 Genomes data to UK10K data increased the number of variants imputed but not imputation accuracy. For variants with MAF < 0.1%, 0.5% to 0.9% of additional variants were imputed. The optimal approach (two reference panels, re-phased, all haplotypes sampled) is 0.531, 0.634, 0.709, 0.771, 0.856, compared to 0.253, 0.338, 0.511, 0.634, 0.780 by using the standard approach (1000 Genomes data alone, phased as-is, 500 haplotypes sampled).

In summary, we demonstrated improved imputation for European samples by using UK10K data and a new strategy to re-phasing and combining WGS reference panels.

Next generation sequencing technologies have rapidly advanced to the point that patient-specific whole-genome sequencing may soon be as routine as X-rays and cholesterol testing. However, the challenge in interpreting the vast amount of data generated by genomic sequencing and effectively using it to guide decisions about an individual’s health care is significant. As part of Mayo Clinic’s Center of Individualized Medicine we are developing a pipeline to analyze whole exome sequencing (WES) of small to large pedigrees and provide insights in genetic disease. In this abstract we focus on the challenge of SNV/SNP selection necessary to perform identity by descent (IBD) analysis for validating the correct relatedness between relatives. For SNV selection we selected samples with genotype data from both WES and SNP chip arrays to select sets of informative markers based on minor allele frequency and linkage disequilibrium. We used the maximum likelihood method available in PLINK and the maximum likelihood method available in PREST for IBD analysis and applied, them to pedigrees of different size. To determine possible bias in the computed IBD, we varied the number of related subjects, the type of relationship and number of variants used in the analysis. We observed that the IBD calculation in PREST is not as susceptible to biased results for close relationship as in PLINK, and including a sample of unrelated subjects helps stabilize the IBD calculation in PLINK and PREST.

A Weighted U statistic for Genetic Association Analyses of Sequencing Data. C. Wu, M. Li, Z. He, Q. Lu. 1) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI; 2) Division of Biostatistics, Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, Arkansas; 3) Department of Biostatistics, University of Michigan, Ann Arbor, Michigan.

Despite the recent success of genome-wide association studies, a large proportion of genetic variants predisposing to complex diseases remain uncovered. Evidence from genetic studies and evolutionary theory has suggested rare variants could play an important role in the biological pathways of complex diseases. The advance of next generation sequencing technology facilitate the generation of massive amount of genetic variants and offers great opportunity to investigate the role of millions rare variants in the genetic etiology of complex disease. Nevertheless, great challenge has also been posed to statistical analyses of high-dimensional sequencing data. The association analyses based on traditional statistical methods endure substantial power loss because of low frequency of genetic variants and extremely high dimensionality of the data. We therefore developed a weighted U statistic for high-dimensional association analysis of next-generation sequencing data. Based on the non-parametric U statistic, our method makes no assumption of the underlying disease model and can be applied to various types of phenotypes (e.g., binary and continuous phenotypes). Through simulation studies and an empirical study, we found our method outperformed a commonly used SKAT method when the underlying assumption is violated (e.g., the phenotype follows a heavily skewed distribution) and attained comparable performance to SKAT when underlying assumption is satisfied. In an empirical study of Dallas Heart Study (DHS) sequencing data, our method was also able to detect the association of ANGPTL4 with very low density lipoprotein cholesterol.

Accurate Genotype Calling with Contaminated Sequencing Data. M. Flickinger, G. Jun, G. R. Abecasis, M. Boehnke, H. M. Kang. Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI.

Advances in next-generation sequencing have enabled a wide range of large-scale genetic studies. While the quality of the sequence data is generally improving, protocols are not perfect, and inevitably some handling errors may occur. One common error is sample contamination through the mixing of two or more samples. We previously developed methods to detect contaminated samples (Jun et al. AJHG 91:839-848, 2013), so that these could be flagged and dropped from analysis. Here we describe new mixture-model-based methods to account for contamination in the genotype likelihoods which result in genotype calls for contaminated samples with accuracy approaching that for uncontaminated samples. To explore the operating characteristics of our method, we simulated contamination in silico by mixing reads from pairs of 198 European 1000 Genomes Project BAM files for whom both high-depth (50-150x) exome and low-depth (4-6x) whole genome sequencing data were available. We then compared the resulting calls to the available Omni 2.5M and Exomechip genotype arrays to evaluate success. In our first experiment, we constructed samples with an 85:15 mix corresponding to 15% contamination. In the high-depth exome data, the overall genotype discordance with the Exomechip decreased from 3.6% to 0.9% by modeling the contamination, compared to 0.3% for datasets with no contamination. For the low-depth data, we used our method to model contamination in the genotype likelihoods and then Beagle for linkage-disequilibrium-aware calling. Here we saw a drop in discordance with the Omni array data from 9.3% to 3.5% compared to 1.1% with no contamination. Furthermore, the false positive and false negative rate (% of incorrect and missed non-reference alleles) were also reduced by >50% for both high- and low-depth data. When the identity of the contaminating sample is known, our approach can incorporate this information resulting in further improvement: reducing discordance in the high-depth data from 0.7% to 0.3% discordance observed when analyzing uncontaminated samples. In conclusion, our methods provide an effective alternative to expensive re-sequencing for accurately genotyping low to moderately contaminated samples.

A Rare Variant Selection Algorithm to Locate Susceptible Rare Variant from Sequencing Data. S. Wang, H. Sun. Dept Biostatistics, Columbia Univ, New York, NY.

Current association methods for sequencing data have been focused on aggregating rare variants across a gene or a genetic region due to the fact that analyzing individual rare variants is underpowered. To identify which rare variants in a gene or a genetic region out of all variants are associated with the outcomes is a natural next step. Here we propose a forward variable selection-based algorithm that is able to identify the locations of potentially susceptible rare variants that are associated with the outcomes with sequencing data. More specifically, we test associations of variants individually in the first step to select the most outcome-related variant that has the maximum test statistic. In step 2, weighted linear combinations of two variants with the selected variant from Step 1 and all other variants one by one are generated and tested to select the most outcome-related two-variant combination. Note that the most outcome-related variant selected in Step 1 could be either risk or protective. In Step 2, we also consider the combination of potentially protective variants through flipping of the variant coding. Similar to a forward variable selection, we keep adding one variant in each step until the test statistic of the linear combination of the combined variants is no longer increasing, and the combined variants are considered as the best combination of potentially causal variants. We evaluated the selection performance of the proposed algorithm through simulation studies where different effect sizes, sample sizes and directions of the effects of the individual rare variants were considered. The results demonstrated that the proposed algorithm is able to select subsets with most of the outcome related rare variants in all simulation scenarios considered. The proposed method was also applied to sequenced data on the ANGPTL gene family from the Dallas Heart Study (DHS). The proposed method was implemented in an R package RVsel which will be freely downloadable at http://www.columbia.edu/~sw2206.
An advantage of Tango's method is that it is rapid to compute, and when tests have been developed, one of which is based on the popular Kulldorff variants in, and around, a gene would benefit genetic association analyses, posed T-sigma-MidP method is recommended for the easiness of its implementation. Studies where oxygen saturation of 767 individuals with 795,736 SNPs were measured for 380 minutes on average. With the development of next-generation sequencing technology, there is a great demand for powerful methods to detect rare causal variants (minor allele frequency MAF < 1%). Testing for each variant site individually is known to be underpowered, and many methods have been proposed to test for the association of a chromosomal region containing multiple variant sites with phenotypes. However, this pooling strategy inevitably leads to the inclusion of a large proportion of neutral variants, which may compromise power. Here, we have proposed a T-sigma-MidP method that combines per-site P-values with weights based on MAFs. Before combination, we first imposed a truncation threshold upon the per-site P-values to guard against noise caused by the inclusion of neutral variants. Simulations have shown that T-sigma-MidP outperforms competing tests in the majority of simulation settings. We applied T-sigma-MidP to the Dallas Heart Study data set, and obtained more significant results in testing for the association of the ANGPTL4 gene with triglyceride and very-low-density lipoprotein. The proposed T-sigma-MidP method is recommended for the easiness of its implementation, its satisfactory power to detect rare causal variants (population MAF < 1%) and/or uncommon causal variants (1% < population MAF < 3%), and its robustness to noise from neutral variants.
A DNA variant caller adapted to assess mitochondrial DNA variation in lymphocytes from 2,000 Sardinians. J. Ding1, C. Sidore1-2, O. Meireles1, M.K. Wing2, F. Busonero3, R. Nagaraja1, F. Cucca3, G.R. Abecasis1, D. Schlessinger1. 1) Laboratory of Genetics, National Institute on Aging, Baltimore, MD; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy.

The degree to which mitochondrial DNA (mtDNA) varies heritably and somatically has been much discussed, but has not been systematically analyzed on a population basis. Further, accumulation of mutations in mitochondrial DNA has been suggested to play an important role in aging. To approach such an analysis of mtDNA variation, genotype calling and analytic programs developed for nuclear DNA must be modified, because each cell has 100-10,000 mtDNA copies that can vary at any site (i.e., heteroplasmasy). We have developed an algorithm that is adapted to identify variants in mtDNA; it incorporates the sequencing error rate at each base in likelihood calculations and is flexible to allow for different allele fractions at a variant site across individuals. Our analysis procedure is also adapted to the circular mitochondrial genome, a key difference from the linear chromosomes assumed by most read mapping algorithms.

We assess homoplasmy and heteroplasmy in mtDNA sequences of lymphocytes from ~2,000 SardinIA Project participants. The distribution of the number of homoplasmy per individual is bimodal, with an average value of 22. The number of homoplasmy with a minor allele fraction threshold of 4% varies considerably among individuals, with most individuals displaying 3 or less homoplasmy, but some showing considerably more. Of the overall heteroplasmy increases with age, but the slope is small, yielding an average increase of ~1 heteroplasmy between ages 20 and 80. As expected, mothers and their children share essentially all homoplasmies but a lesser proportion of heteroplasmy. Overall, it is clear that mitochondrial variants do accumulate in lymphocyte populations as a function of age, but the extent of accumulation remains hard to assess accurately because many variants are on par with sequencing error rate. Currently, attempts are being made to increase resolving power by using higher sequence coverage and by developing new analytic methods.

The results to date provide information about mtDNA haplogroups and the inheritance of homo- and heteroplasmy in Sardinia. The algorithm can be further extended in several ways: for example, to study mtDNA from cloned normal cells in greater depth, and to investigate the nuclear DNA variability in cancer cells.

Understanding the limits of pooled next-generation sequencing to identify causal modifier variants in Cystic Fibrosis. J. Gong1, F. Lin3, T. Chiang2, M. Keenan1, M. Mittl1, D. Scoate1,4, W. Li1,4, L. Sun1,5, R. Nagaraja2,6, K. Keenan1, T. Rommens1,2, L. Strug1,4. 1) Program in Child Health Evaluation Sciences, the Hospital for Sick Children, Toronto, ON, Canada; 2) Program in Genetics and Genomic Medicine, the Hospital for Sick Children, Toronto, Ontario, Canada; 3) Program in Physiology and Experimental Medicine, the Hospital for Sick Children, Toronto, Ontario, Canada; 4) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 5) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 6) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

Modifier genes contribute to variability in morbidity across multiple organs affected in cystic fibrosis (CF). SLC26A9, SLC9A3 and SLC6A14, were shown to contribute to CF intestinal obstruction (IO) in a hypothesis driven genome-wide association study (GWAS) by Sun et al. (2012) (Nat Genet 44:562-9). We performed enrichment analysis on the 3 genes from 200 CF individuals, 100 with MI cases, and identified potential modifier variants in SLC9A3 and SLC6A14 by whole genome sequencing (WGS). We compared these to previously identified variants identified by pNGS enrichment of these genes and imputed SNPs, resulting in high concordance between the variants identified by the two methods. Further, we performed Sanger sequencing in one region of SLC6A14 selected by the association results to assess the reliability of our findings. Our experience suggested that pNGS identified all the variants detected by SS in this region. However, the local MAF estimates from both pNGS and imputation differed substantially from that provided by SS, resulting in discrepant association interpretation. Our experience suggested that pNGS is reliable for variant detection. Its estimation of MAF was reasonable on average, but a sub-gen region was observed where the MAF estimates (and thus association results) differed substantially from values obtained from SS. The discrepancies were functions of region-specific qualities of the sequence data and genetic architecture, neither of which can be known a priori. There was no general agreement among the tested pNGS association methods, limiting the usage of pNGS data for reliable association testing.

Accurate local ancestry inference in exome sequenced admixed individuals using off target sequence reads. Y. Hu1,2, C. Willer1, X. Zhan2, H. Kang3, G. Abecasis1,2. 1) Integrative Biology, University of California, Berkeley, CA, 94720, US; 2) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48109, US; 3) Department of Human Genetics, University of Michigan, School of Medicine, Ann Arbor, MI, 48109, US.

Estimates of the ancestry of specific chromosomal regions in admixed individuals are useful for studies of human evolutionary history and disease gene mapping. Previously, this ancestry inference relied on high quality genotypes from genome-wide association study (GWAS) arrays. These high quality genotypes are not always available when samples are exome-sequenced, and the strategy of choice for many ongoing genetic studies. Here we show that off-target reads generated during exome sequencing experiments can be combined with on-target reads to accurately estimate the ancestry of each chromosomal segment in an admixed individual. To reconstruct local ancestry, our method SEGMIIX models aligned bases directly instead of relying on hard genotype calls. We evaluate the accuracy of our method through simulations and analysis of samples sequenced by the 1,000 Genomes Project and the NHLBI Grand Opportunity Exome Sequencing Project. In African-Americans, we show local ancestry estimates derived using our method are very similar to those derived using Illumina’s Omni 2.5M genotyping array, and much improved in relation to estimates that use only exome genotypes and ignore off-target sequencing reads. Software implementing this method, SEGMIIX, can be applied to analysis of human population history or for disease gene mapping studies in admixed individuals.
1891F Flexible and robust methods for rare-variant testing of quantitative traits in pedigrees. Y. Jiang1,2, K.N. Conneely1,2, M.P. Epstein1,2, 1) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA.

Rare-variant sequencing studies are increasingly popular strategies for investigating the missing heritability of complex human traits. Although a series of statistical methods have been developed to analyze rare variants, the majority of tests are restricted to case-control and population-based study designs. Few statistical methods have been developed to analyze rare variants in family-based studies. As more and more resequencing studies employ family-based designs to overcome potential bias caused by population stratification and to study cosegregation patterns of causal variants, the development of powerful methods for family-based analyses of rare variants is needed.

With this in mind, we propose a rare-variant association test for quantitative traits in families that uses a kernel framework. Within a region of interest, the model partitions a family member’s genotype at a rare variant into a within-family component (robust to population stratification but sensitive to genotype error) and an orthogonal between-family component (sensitive to population stratification but less sensitive to genotype error). For large-scale sequence data, our approach first constructs a kernel test using the between-family component as a screening tool to identify top hits. We then filter these top hits using a kernel test based on the robust and independent within-family component. Unlike other methods, which assume that all variants in the tested region have the same direction of effect on the phenotype, our method remains powerful when causal variants in the test region may have differing direction of effect. Our method can also accommodate covariates and exhibits in that it can still perform well under partial missingness of parental genotypes. Finally, our method has the practical benefit of permitting efficient calculation of p-values based on asymptotics rather than requiring computationally-expensive permutations, which enables practical application to genome-wide data. Using simulated data, we have already shown that our method can avoid inflated false positive rates caused by population stratification. We will also use simulated data to explore power and further apply our method to the Sardinia sequencing dataset available through dbGaP for illustration purposes.


GWAS aims to find SNPs associated with phenotypes of interest. Since detected SNPs are likely to be in linkage disequilibrium with the truly causal SNP, rather than being causal themselves, a common follow-up study will be to sequence an area around one or more regions containing such SNPs. Multiple non-trivial questions arise: What regions should we choose? How wide should those regions be? How deep should the coverage be? What study design should we use: two-phase case-control (Schaid et al., Genet Epidemiol 2013;37: 229-38.), family-based (Xu & Zhi, AJHG 2012;90:1028-45), etc.? How do we best analyze the resulting data? We will describe our efforts as part of NIH’s “GWASeq” consortium, which aims to provide the empirical data that will allow the community to answer these questions. Our component of the consortium focuses on data from the Colonrectal Cancer Family Registry (C-CFR) (Newcomb et al., CEBP 2007;16: 2331-43). We have sequenced around 4000 samples from the C-CFR, across 10 regions, at more than 50X coverage. Our data is a mixture of population-based and pedigree-based samples. We will describe the study, the data that resulted from it, and the conclusions that can be drawn regarding design of such studies in future and the implications regarding so-called “missing heritability.”

1893T Are base qualities necessary in the context of high coverage sequencing? M.W. Snyder, J. Shendure. Genome Sciences, University of Washingon, Seattle, WA.

Base quality scores from next-generation sequencing instruments are used in a variety of data processing steps. Some aligners use quality scores to guide read placement on a reference. During variant discovery and variant calling, probabilistic and likelihood-based models may make use of quality scores in genotype inference. Downstream, post-processing steps often include filtering variants on the basis of base quality score metrics. However, the storage and use of these quality scores has associated costs. First, intermediate file sizes are significantly increased for reads in the FASTQ format, the stored base qualities nearly double the overall size of each file. These same qualities may then be propagated forward to SAM or BAM format, where file size is again inflated by the presence of base quality information. Additionally, the use of base qualities during likelihood calculations, read placement, or variant filtering imposes a computational penalty on data processing pipelines. Conventional wisdom suggests that these costs are justified by the concomitant increase in genotype accuracy to which they give rise. Here, we challenge the validity of this assumption in the context of very high coverage datasets generated on modern sequencing instruments. We attempt to use the counting statistics naturally arising from high coverage sequencing as surrogates for base qualities, and investigate the impact on genotype accuracy of processing sequencing data that has been stripped of base quality scores. To wit, we perform alignment, variant discovery, and genotyping without base quality scores, using a combination of conventional and custom software tools, and compare our results to variant calls from existing data analysis pipelines and standard genotypes. We apply our approach to a family trio (NA12877, NA12878, and NA12882) sequenced to 200X coverage on an Illumina HiSeq 2000 system and for whom gold standard genotypes are available from orthogonal methods. In preliminary investigations, we achieve approximately 99% non-reference genotype concordance with array-based genotyping methodologies, with the original reads and evaluate genotype accuracy as a function of mean coverage. We additionally describe the cost savings in terms of reduced computational time and smaller storage footprint achieved through the application of this approach.

1894F A UNIFIED SEQUENCE KERNEL ASSOCIATION TEST ALLOWING FOR ADMIXED SUBJECTS WITH ARBITRARY RELATEDNESS. W. Ouyang1,2, H. Deng1,2, H. Qin1,2. 1) Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, New Orleans, LA 70112; 2) Center for Bioinformatics and Genomics, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, New Orleans, LA 70112.

Rare genetic variants may potentially contribute to more missing heritability complex human diseases. Over the recent years, next generation sequencing studies (NGSs) have been conducted to provide more comprehensive and accurate description of rare and common variants, and a number of statistical methods have been developed for identifying sequence associations. Current prominent sequence association methods assume homogeneity and/or un-relatedness of study subjects. Such assumptions could be severely violated in the samples from admixed populations, e.g., African Americans and Hispanic Americans. Population stratification and cryptic relatedness are two important confounders in genetic association studies in admixed populations. In this paper, we have developed a unified sequence association test (AdmSKAT) to allow for the NGS data of admixed subjects with arbitrary pedigree structure and diverse directions of effect of genetic variants on phenotype. Our AdmSKAT approach extends the conventional population based sequence kernel association test (SKAT) and family-based SKAT by jointly modeling population stratification, relatedness and local ancestry-by-genotype interaction to ensure type I error control and to improve statistical power. We simulated extensive sequence data of admixed families for method comparisons. Under the null of no sequence association, our AdmSKAT and FamSKAT control type I error rates, whereas naive application of the conventional SKAT and FLM lead to inflated type I error rates. Our AdmSKAT were more powerful than the FamSKAT when applied to all the simulated scenarios of admixed families. AdmSKAT has higher power than competing method in many different scenarios. We then apply our AdmSKAT to analyze the real and simulated sequence data on hypertension of Mexicans from GAW18.

The CONVERGE study of Major Depression has collected low-coverage (1x) sequencing data on 12,000 Chinese women. This study is one of the largest whole-genome sequencing studies currently underway. To accurately infer genotypes in these samples, LD-based genotype refinement methods are needed. However, the study size makes this a challenging proposition. We have investigated several computationally tractable strategies for calling genotypes. Current MCMC schemes for phasing and genotype calling do not explicitly encapsulate the local IBD structure between individuals. We have investigated several additional MCMC schemes to sample haplotypes and genotypes that underlie each sample that try to learn details about local haplotype sharing as the MCMC sampler evolves. We have also investigated whether current methods can be made to produce acceptable results within practical time frames.

We have applied these methods to genotype likelihoods at ~14 million 1000 Genomes Project Phase 1 (TGP1) SNPs polymorphic in TGP1 Asians. We were able to infer genotypes in the whole CONVERGE cohort at a computational cost equivalent to ~2,000 CPU days on a single core of a current Intel processor. We have measured the accuracy of our imputed genotypes using external validation genotypes on 16 individuals typed on the genome-wide Illumina Zhonghua-8 SNP chip with 890,371 SNPs. We find that the mean imputation r² at MAFs <0.5%, 0.5-1%, 1-2%, 2-5%, and >5% are 0.21, 0.57, 0.70, 0.82, and 0.92. Overall, these results illustrate the feasibility and accuracy of using very low-coverage sequencing in large studies.

Modeling gene expression and rare sequence variation to identify genes and subnetworks underlying autism risk. L. Liu, J. Lei, S. Sanders, J. Wilkay, Y. Kousathanas, C. Klei, C. Liu, X. He, A. Ma’ayan, J. Noonan, N. Sestan, J. Buxbam, M. State, B. Devlin, K. Roeder. 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; 2) Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA; 3) Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 4) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 5) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA; 6) Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; 7) Department of Neurobiology and Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut, USA; 8) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 9) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 10) Program on Neurogenetics, Yale University School of Medicine, New Haven, Connecticut, USA; 11) Child Study Center, Yale University School of Medicine, New Haven, USA; 12) Department of Psychiatry, Yale University School of Medicine, New Haven, USA.

Rare variants, especially de novo loss of function (LoF) mutations, have proven effective at identifying genes affecting risk for autism spectrum disorders (ASD). When a de novo LoF mutation falls in a gene, and more than one ASD proband carries such a mutation, the gene is likely to be a risk gene. Based on the rate of de novo mutations in ASD probands versus their siblings, we infer that roughly half of the genes with LoF mutations observed only once per generation in the combined sample are risk genes. To extract more information from Whole Exome Sequence (WES) data, He et al. develop a statistical model that integrates data from family and case-control studies to infer the likelihood a gene affects risk for ASD; this model, called TADA for Transmission And De novo Association, summarizes genetic evidence for a gene affects risk for ASD. Still, given the limited WES data, can we garner more information regarding ASD risk? Relative to genes with strong genetic support for involvement in ASD, we hypothesize that genes expressed at lower frequency in specific brain regions with high gene expression and strong correlated expression, would themselves be more likely to affect risk for ASD because they are functionally interrelated. To find these sub-networks of co-expressed genes we jointly model two kinds of data: gene co-expression and genetic data. We model the dependency in the network, which is inherent in the correlation of a gene’s state, ASD risk gene or not, with its near neighbors states. The statistical analysis identifies over a hundred genes that plausibly affect risk, a large fraction of which are not known to affect risk for ASD. For those that are novel, many were implicated on the basis of the statistical model, not the strength of the genetic evidence alone. A third of the genes with a single LoF de novo mutation were implicated. These results expand our understanding of the neurobiology of ASD.
Gene-based Generalized Functional Linear Models for Case-Control Association Studies. R. Fan¹, Y. Wang¹, J.L. Mills², A.F. Wilson³, J.E. Bailey-Wilson⁴, M.M. Xiong⁵. 1) Biostatistics and Bioinformatics Branch, Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Rockville, MD; 2) Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; 3) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA.

Gene-based association testing problems involve tens of thousands of hypothesis tests and it is well known that traditional family-wise error rate control will be too conservative for large multiple testing problems. Standard false discovery rate (FDR) control procedure improves power, but because this approach focuses solely on the p-values of individual hypotheses, this approach too can have suboptimal power by ignoring correlations between individual hypotheses. When substantial prior knowledge is available, however, power can be increased if the prior information and the data can be modeled appropriately. We propose a hierarchical hidden Markov random field (HMRF) model for gene-based association using a graph built from gene expression data. The prior information from the network can be applied as a reference for specific disease association studies.

Pool aliquots shows an average read depth (RD) of 122x and 116x for 200- and 300-PoolA and 200- and PoolB, respectively, yielding a total of 59,010 on-target variants. The allele counts were combined for the 2 aliquots and frequencies were compared to array data for 1574 SNPs previously genotyped for the individuals in this pool, revealing an accuracy of r²>0.90 at an on-target RD <100x and r²>0.96 at RD ≥100x. The 59,010 variants were also filtered against the Online Mendelian Inheritance in Man (OMIM) database yielding 278 (0.47%) mutations, of which 61 of the OMM variants were classified as pathogenic mutations. Among the mutations previously described in this population these were those for hereditary hemochromatosis (HFE p.H63D), Factor V Leiden (F5 p.R534Q), and Canavan disease (ASPA p.Y231X). Ameliorate addition, because the newly identified conditions for which mutations were identified were autosomal recessive hypoparathyroidism (PTH p.R83X), Fletcher factor deficiency (KLKB1 p.S143N), autosomal dominant deafness (MYO1A p.G662E), Warfarin sensitivity (CYP2C9 p.I359L), trimethylaminuria (FM03 p.V257M), hyperglycinuria (SLC6A20 p.1198M), ataxia spinocerebellar ataxia type 2 (ATSCA2 p.P684A), and tubular acidosis (NPPA p.M580T). Pooled exome sequencing provides a cost effective method for establishing population mutation profiles, such as the profile developed here for the Ashkenazi Jewish population, and if linear will be an important reference for specific disease association studies.
1902T

A New Statistical Method for Identifying Differentially Methylated Regions in Complex Diseases. P. Liu1,2, C. Chen1,2, X. Hua1,2, Y. Lu1,2, M. Liang1. 1) Department of Physiology, Medical College of Wisconsin, Milwaukee, WI; 2) Cancer Center, Medical College of Wisconsin, Milwaukee, WI.

DNA methylation is an important epigenetic modification that regulates transcriptional expression in complex diseases, such as cancer. It occurs when a methyl group becomes covalently attached to the 5-carbon position of a cytosine residue at a ‘CpG’ site. Of particular interest are CpG sites located in CpG islands which are clusters of CpG sites showing dynamic variations of methylation levels and often colocalized with cis regulatory elements such as gene promoters. Increased DNA methylation in CpG islands within or close to gene promoters is typically associated with transcriptional repression or gene silencing. Genomic segments with enriched 5-Methylcytosine marks can be mapped to single-base resolution using next-generation sequencing. Such high-throughput base-resolution methylation data have unique features and hence require the development of new analytical approaches. These important features include dependence of methylation among CpG sites within a CpG island and the influence of depth of sequence coverage on measuring methylation rate at CpG sites. Here, we proposed a new statistical method for identifying differentially methylated CpG regions from base-resolution methylation sequencing Data. Our approach accounts for correlation structures among CpG sites and incorporates depth of sequence coverage as weights for measuring methylation rate in CpG sites. We evaluated the performance of our new approach and compared it to other methods using the most commonly-used t-test and Fisher’s exact through a wide range of simulations. The simulation results show that regular tests for analyzing methylation sequencing data dramatically inflate type I error rates in the presence of physical proximity of CpG sites. While our new method is a valid test and gains statistical powers when weighting methylation rate by depth of sequence coverage, it will be much more powerful when incorporating additional information such as CpG island and gene promoter proximity. Our analysis of methylation sequencing data from non-small cell lung cancers and revealed novel methylation patterns and statistical inferences, which can be incorporated as covariates into a logistic regression model. We then merge the triads into a network. We apply community detection algorithms in order to identify homogeneous subgroups or communities, which can further be incorporated as covariates into a logistic regression model. We apply our method to populations from different continents in the 1000 Genomes Project and evaluate the type I error based on the empirical p-values. Our simulation results suggest that the network approach provides a more precise information of population structure than existing methods.

1903F

A nonparametric model for haplotypes in population bottlenecks. L.T. Elliott1,2, Y.W. Teh1,2, 1) Gatsby Computational Neuroscience Unit, University College London, London, UK; 2) Department of Statistics, University of Oxford, Oxford, UK.

Clustering methods based on hidden Markov models are useful for approximating the haplotype structure of genetic sequence data that have undergone a recent population bottleneck or admixture. In these models (such as fastPHASE, IMPUTE2 and BEAGLE), latent cluster indicators can specify the genetic founder or component from an admixed population from which a loci on a given sequence originates. In this work we present BNPPHASE (Bayesian nonparametric Phase), a nonparametric model of haplotype structure based on a hierarchical Dirichlet process hidden Markov model (HDP-HMM). Bayesian nonparametric models handle latent variable domains with unbounded dimensionality and in BNPPHASE the number of genetic founders is unknown but inferred concurrently with the rest of the model without a separate model selection phase. BNPPHASE describes genetic sequences in which the proportions describing the distribution of each one of an unknown number of genetic founders varies along the chromosome reflecting ancestral recombination events. BNPPHASE advances previous work in nonparametric genetic models by providing nonhomogeneous cluster proportions and by identifying haplotypes across loci through the use of the recently proposed ‘sticky’ HDP-HMM leading to more efficient inference and to posterior distributions that more closely resemble the genetic process. The popular fastPHASE model can be seen as a finite truncation of BNPPHASE. The BNPPHASE model can be described by a generative process in which the cluster indicators for each sequence either remain the same from one loci to the next or, with a probability given by a recombination rate, are updated according to a standard hierarchical Dirichlet process model. In a series of experiments involving simulated data simulated from a population bottleneck with K ancestral genetic founders, we hold out data and compare the imputation accuracy of BNPPHASE with that of its finite truncations. We show that the flexibility of BNPPHASE, which stems from its nonparametric nature, allows BNPPHASE to achieve higher imputation accuracy even after the capacity of a given finite model is exhausted. Furthermore, we show that BNPPHASE can be used to estimate the number of genetic founders in a sample and that it can discover situations in which the total number of ancestral genetic founders changes along the chromosome due to selection or genetic drift.

1904W

Using network methodology to infer population substructure. D. Prokopenko1, C. Lange1,2,3,4, J. Heckler1, P. Costa1, E.K. Silverman5, H. Loehlein Fier1. 1) Institute of Genomic Mathematics, University of Bonn, Bonn, Germany; 2) Department of Biostatistics, Harvard School of Public Health, Boston, United States; 3) Channing Laboratory, Brigham and Women’s Hospital, Boston, United States; 4) German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany.

One of the main caveats of association studies is the possible affection by bias due to population stratification. Existing methods rely on model-based approaches like structure (Pritchard et al. 2000) and ADMIXTURE(Alexander et al. 2009) or on principal component analysis like EIGENSTRAT(Price et al. 2006, Patterson et al. 2006). Here we describe the problem of population substructure from the graph-theoretical point of view. We group the sequenced individuals into triads, which depict the relational structure, utilizing a similarity measure, i.e. covariance matrix. We then merge the triads into a network. We apply community detection algorithms in order to identify homogeneous subgroups or communities, which can further be incorporated as covariates into a logistic regression model. We apply our method to populations from different continents in the 1000 Genomes Project and evaluate the type I error based on the empirical p-values. Our simulation results suggest that the network approach provides a more precise information of population structure than existing methods.

1905T

eqtl-tools: a toolkit for scalable eQTL mapping. A. Di Nardo1,2, K. Hao1,2, 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Institute of Genomic and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY.

Motivation: Expression Quantitative Trait Loci (eQTL) mapping on modern datasets requires the evaluation of a massive number of correlations. Current tools are either too slow or have hard limits on the input/output data formats and memory requirements. They also lack the flexibility to accommodate sex chromosomes or nonparametric models. Implementation: eqtl-tools is a set of command line utilities for fast and comprehensive eQTL mapping which uses plain text, tab-delimited tables as the main data format, allows for easy control of memory usage, and handles missing data consistently. Conclusions: eqtl-tools significantly lowers the amount of RAM necessary to perform eQTL analysis, while incurring in moderate speed penalty w.r.t. the best alternative solution to date. It handles chromosome X and facilitates integration into custom data analysis pipelines. It also allows more sophisticated modeling, e.g. nonparametric methods, and calibrates empirical false discovery rate (FDR) through permutation. Availability: The package is easy to deploy, self-documenting, and released under the GPL-v3 license on eqtl-tools.googlecode.com. Contact: antonio.dinarzo@mssm.edu.
provide increased sensitivity with respect to traditional gene expression studies.

We first combined studies within the same ancestral group: African (LWK, MKK, YRI), East Asian (CHB, JPT) and European, South and Hispanic (CEU, GIH, MEX). Subsequently, we combined the 'ancestry-specific' association statistics across ethnicities. With this data, we aimed to evaluate the improvement in discovery and fine-mapping of eQTLs through trans-ethnic meta-analysis using 1000 Genomes imputation. After imputation, the lead SNP at 1,629 probes achieved genome-wide significance (log10 BF > 6), compared with 1,591 for HapMap genotypes alone. An improvement in the strength of the association at the lead SNPs was observed after imputation (Probe ILMN_10409_6860870 (chr10-105156223); HGNCS: USMGS; ICGS: 1,187 SNPs; growth 5) has the greatest increase in association signal after imputation. The lead typed SNP is rs7831 (log10 BF = 132), whilst the lead imputed SNP is 10-105161778 (log10 BF = 250). To determine which ancestry groups are most informative for fine-mapping, we defined 99% credible sets of imputed and HapMap SNPs that are most likely to contain the causal variant. As expected, the African group has the lowest median credible set size (44 SNPs, 154 Kb), followed by the European, South Asian and Hispanic group (20 SNPs, 1072 Kb), and East Asian group (SNP, 719 Kb). Considering all ancestral groups leads to further improvements in resolution (median 99% credible set size of 8 SNPs over 19 Kb). This study has demonstrated that trans-ethnic meta-analysis and imputation provides a powerful means for discovery and fine-mapping of eQTLs.

Trans-ethnic meta-analysis can increase power to detect expression quantitative trait loci (eQTLs) and improve fine-mapping. We considered lymphoblastoid cell line expression in Phase 3 HapMap populations from multiple ancestry groups: CEU (CEU, European), Chinese (CHB, East Asian), Gujarati Indians (GIH, South Asian), Japanese (JPT, East Asian), Luhya (LK, African, Kenya, East African), Maasai (MKK, African, Kenya, East African), Yoruba (YRI, African). Genotypes were available for all samples and were used as a scaffold for imputation with the Phase 1 1000 Genomes panel (June 2011). We tested for association with expression of 1639 probes within each study for cis-LOH (1 Mb upstream and downstream of each gene) and then combined studies via trans-ethnic meta-analysis implemented in MANTRA, a Bayesian approach which allows for heterogeneity in allelic effects between groups. We first combined studies within the same ancestral group: African (LWK, MKK, YRI), East Asian (CHB, JPT) and European, South and Hispanic (CEU, GIH, MEX). Subsequently, we combined the 'ancestry-specific' association statistics across ethnicities. With this data, we aimed to evaluate the improvement in discovery and fine-mapping of eQTLs through trans-ethnic meta-analysis using 1000 Genomes imputation. After imputation, the lead SNP at 1,629 probes achieved genome-wide significance (log10 BF > 6), compared with 1,591 for HapMap genotypes alone. An improvement in the strength of the association at the lead SNPs was observed after imputation (Probe ILMN_10409_6860870 (chr10-105156223); HGNCS: USMGS; ICGS: 1,187 SNPs; growth 5) has the greatest increase in association signal after imputation. The lead typed SNP is rs7831 (log10 BF = 132), whilst the lead imputed SNP is 10-105161778 (log10 BF = 250). To determine which ancestry groups are most informative for fine-mapping, we defined 99% credible sets of imputed and HapMap SNPs that are most likely to contain the causal variant. As expected, the African group has the lowest median credible set size (44 SNPs, 154 Kb), followed by the European, South Asian and Hispanic group (20 SNPs, 1072 Kb), and East Asian group (SNP, 719 Kb). Considering all ancestral groups leads to further improvements in resolution (median 99% credible set size of 8 SNPs over 19 Kb). This study has demonstrated that trans-ethnic meta-analysis and imputation provides a powerful means for discovery and fine-mapping of eQTLs.
1910W SKAT Admix: extending the SKAT method for rare variant association to admixed populations. A.E. Byrnes1, M.C. Wu1, M. LP, Y. Li1,2,4. 1 Biostatistics, University North Carolina, Chapel Hill, NC; 2 Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia Pennsylvania; 3 Genetics, University of North Carolina, Chapel Hill, NC; 4 Computer Science, University of North Carolina, Chapel Hill, NC. The Sequence Kernel Association Test (SKAT), first proposed in 2011, was designed to measure association between a set of genetic variants (including rare variants) and complex genetic traits quickly and efficiently. Here, we expand the SKAT methodology for use in admixed populations by first predicting the ancestry for each individual, at each locus, by utilizing additional sequencing data from each of the parent populations. We then apply the SKAT methodology to minor alleles likely to be inherited from each parent population separately, resulting in a SKAT statistic for each parental population. For example, in the case of African Americans, we conduct one SKAT test for those markers of European ancestry and one for those of African ancestry. To evaluate this method, we present a series of simulations intended to mimic large sequencing studies of a quantitative trait in African Americans. In our simulations, the chosen variants are only causal if they originate from one of the parental populations, either African or European. We also examine different numbers of causal variants and directions of association. We find that SKAT Admix has increased power to find association compared to SKAT alone in most cases. In addition, the SKAT method has been proven to preserve type I error without requiring permutation. In our simulations, SKAT Admix also preserves Type I error for variants of both African and European ancestry. Data from the 1000 Genomes Project suggest that many more rare variants are population specific, as compared to common variants. It has also been suggested that rare variants show stronger patterns of stratification than common variants. For these reasons, we suggest a test that attempts to separate genomic data based on the predicted population of origin for admixed data.

1911T The visualization of probabilistic results from consumer genetic testing for ethnicity at AncestryDNA. R.E. Curtis, K.H. Freestone, M.J. Barber, J.M. Callaway, K. Noto, Y. Wang, C.A. Ball, K.G. Chahine. AncestryDNA, Provo, US. An important, but often overlooked, challenge in consumer genetics is the design of engaging and informative data visualization strategies that help consumers understand and fully appreciate the results of their genetic tests. Successful data visualizations that allow accurate and meaningful interpretation of consumer genetic test results must (1) help consumers overcome erroneous preconceptions or assumptions, (2) bridge gaps in the genetics education of consumers and (3) communicate the uncertainty associated with probabilistic predictions to consumers who may not have a strong understanding of statistics. Conveying uncertainty associated with predictions is crucial because many aspects of genetic science rely on probabilistic theory, including identity by descent (i.e., cousin matching), admixture predictions (i.e., ethnicity), and relative disease risk. If the stochastic nature of genetic algorithms is not properly conveyed in the visualization, naive users often will either whole-heartedly accept the results as ground truth or dismiss them altogether. At AncestryDNA, we have considered this problem in the context of an admixture prediction based on autosomal SNP testing. Over the past year, we have delivered over 125,000 admixture predictions to consumers across the US and conducted multiple surveys to assess our customer's satisfaction with and comprehension of those predictions. User surveys revealed several areas for improvement: some consumers had incorrect assumptions about the test; many consumers were confused about genetic concepts and most consumers were unable to understand the probabilistic nature of the predictions. In response to our consumer reactions, we have developed new visualization tools that attempt to address the types of shortcoming described above and to create consumer-friendly approaches to describe statistical concepts such as confidence intervals, sensitivity, positive predictive value, and standard deviations. As part of the natural life cycle of iterative consumer-facing improvements, we survey and evaluate our user interfaces using systematic questionnaires and focus groups to judge which tools most effectively convey the concept of genetics and probabilistic theory to a largely naïve consumer population. We will present our findings based on the conviction that the principles we have applied in our iterative development, refinement, and user experience can also extend to other aspects of consumer genetics.

1912F Next Generation of Genotype Imputation Methods. S. Das, G.R. Abe- cassis. Department of Biostatistics, University of Michigan, Ann Arbor, MI. Genotype imputation is a key step in the analysis of human genetic studies as it facilitates in increasing power of gene mapping, enables combination of results across different studies and accelerates fine-mapping efforts. Imputation works by finding the shared haplotype segments between the sample individuals typed on a custom SNP array and a reference panel of more densely typed individuals (e.g. The International HapMap Project, The 1000 Genomes Project etc.). Advances in high-throughput sequencing technologies have resulted in rapid increase in the size of these publicly available data-sets. Using these as reference panels would soon result in prohibitive computational costs. We introduce a strategy called ‘state space reduction’ which reduces the description of short genomic regions to the number of distinct haplotypes rather than the total number of haplotypes. The existing formulas have been refined to work with only these distinct haplotypes in a series of short genomic regions which covers the whole genome. The proposed algorithm maintains the accuracy of the current methods while reducing the computational costs. We also formulate a recursive algorithm to find the optimal structure/design of such genomic regions.

1913T SECA: SNP error concordance analysis using genome-wide association summary results. D. Nyholt. Neurogenetics Lab, QIMR, Brisbane, QLD, Australia. Epidemiological and clinical studies indicate many human complex disorders co-occur within an individual, while family and twin studies suggest correlations in familial and genetic liabilities. To date, over 1350 genome-wide association studies (GWAS) have been performed to identify common single nucleotide polymorphisms (SNPs) associated with disease endpoints or quantitative traits. A surprising finding of GWAS is that many loci show pleiotropic effects by being associated with more than one distinct phenotype. Indeed, a study of 1380 genes and 1687 SNPs listed in the NHGRI Catalog of Published GWAS (February 4, 2011) found 16.9% of genes and 4.6% of SNPs show pleiotropic effects (Sivakumar et al. 2011). Identifying and taking advantage of polygenic overlap across phenotypes can improve detection of genetic risk factors, because when risk is correlated across phenotypes, pooled analyses will be better powered than individual-disorder analyses.

To facilitate the discovery of pleiotropic effects and examine polygenic risk shared across two phenotypes, I have developed a user-friendly web-based application called SECA to perform SNP effect concordance analysis. SECA takes two input files of GWAS summary results, each containing: i) SNP rsID, ii) effect allele (EA), iii) non-effect allele (NEA), iv) p-value from association test, and v) regression coefficient (beta), odds ratio, or z-score for the EA relative to the NEA. SECA first aligns the SNP effects across the two GWAS summary results (dataset1 and dataset2), and extracts a subset of independent SNPs (randomly or prioritized by p-values in dataset1) via linkage disequilibrium clumping. Next, SECA tests for enrichment of overlapping (pleiotropic) SNPs and whether the effect directions are concordant across the datasets via binomial, Fisher’s, and false discovery rate statistical tests of dataset SNPs, after conditioning on association results in dataset1. SECA also prepares Q-Q, true discovery rate and bar plots to visualize enrichment overlap.

Using publicly available GWAS summary data, SECA corroborates recent results from the Cross-Disorder Group of the Psychiatric Genomics Consortium (Lancet 2013 381:1371-9), finding significant polygenic overlap between bipolar disorder, major depressive disorder, schizophrenia and autism spectrum disorder, but not with attention deficit-hyperactivity disorder.


The goal of the presentation is to test Hardy-Weinberg Equilibrium for Multi-allelic markers a la Fisher. The test involves looking at all possible genotype data sets with the same allele frequencies. This collection of data sets could be humongous. If this is the case, Fisher’s exact test cannot be implemented. We have developed a Markov base to negotiate the collection. We can then implement a Markov Chain Monte Carlo Simulation Algorithm exploiting the Markov Base to carry out the Fisher’s exact test. An R code will be presented to complete the computations. The code will be run on some examples.
1915F

Background: Genome-wide association studies (GWAS) have identified more than 10,000 associated SNPs with 840 traits. Despite this success, there still remains the problem of ‘missing heritability’ for most traits. One contributing factor may be the result of examining one marker at a time as opposed to a group of markers that are biologically meaningful in aggregate. To address this problem, a variety of gene- and pathway-level methods were developed to identify putative biologically relevant associations. Markers are put into either gene-level units or pathways as designated in databases such as Gene Ontology (GO). A simulation was conducted to systematically assess the performance of gene- and pathway-level methods. Methods: Using genetic data from the Wellcome Trust Case Control Consortium (WTCCC), case-control status was simulated based on an additive polygenic model where cases have more risk alleles. A total of 20 pathways and 226 genes were selected from GO biological processes. We evaluated 20 methods (e.g. VEGAS, MAGENTA, GATES, HYST, SRT, Fishers, etc.) based on their sensitivity, specificity, type I and type II error. The influences of gene and pathway size, number of causal single nucleotide polymorphisms (SNPs) in each gene/pathway, and effect size were assessed. Simulations using two sample sizes were examined: traditionally underpowered (n=250 cases, 250 controls) and larger (n=2250 cases, 2250 controls). Results: Despite low overall sensitivity (20-60%), specificity was high (89-100%) with low type I error (0.1-6%). Classical methods, not developed to handle linkage disequilibrium, have higher sensitivity, but higher type I error. Newer methods that incorporate the achieved correlation in the database were underpowered to detect genes with smaller effect sizes, but type I error was low. All methods were able to detect genes that would have been ignored in a traditional GWAS. Pathway-level methods’ performance was dependent upon database annotation. Gene-level methods can rely on annotation from standardized databases but pathway-level methods use a homogeneous pool. Conclusions: The interpretation of gene- and pathway-level methods is dependent upon their specific, inherent assumptions and the resources they rely on. Low type I error is important for gene-level methods, but pathway- and pathway-level methods may provide valuable insight into the ‘missing heritability’ of traits.

1917T
Accurate and rapid genetic analysis of genomic data under mixed model with multiple variance components. H.M. Kang, S. Yang. Biostatistics Dept, Univ Michigan, Ann Arbor, Ann Arbor, MI.

Ultra-high dimensional genetic analysis of genomic data, such as expression quantitative loci (eQTL) mapping, becomes increasing important to unravel the regulatory mechanisms of genetic variation causing molecular and phenotypic changes. Analysis of high throughput sequencing data dramatically increased the hypothesis space and the computational burden. At the same time, the inherent confounding factors such as batch effects or population structure requires us to apply sophisticated statistical models such as mixed models, which even further increase the computational costs compared to simple linear models. Here we demonstrate an accurate and computationally efficient procedure for genetic association analysis with high dimensional expression data under linear mixed models with multiple variance components, at a computational cost comparable to the simple linear model. Our methods have three important advantages over existing methods. First, the computational complexity of our method is linear to the number of samples, markers, and genes, similar to the GRAMMA-Gamma method (Svishecheva et. al. 2012). However, our method does not rely on the simplifying assumption of constant inflation factor across the genome. Instead, it accurately incorporates marker-specific inflation factor (Kang et. al. 2010) without sacrificing the computational complexity. Second, our method robustly deconvolutes the variance component attributed by the technical batch-effects from the variance component attributed by the hidden sample structure using an importance sampling approach. As a result, our method robustly controls Type I error for each phenotype while avoiding overcorrection due to misspecified correlation structure between samples. Third, our method achieves linear time estimation of variance component parameters for multiple variance components by utilizing multiple sets of precomputed singular vectors. As a result, the computational cost of our method across n samples, m markers, and g genes becomes O(nmg) when n is smaller than g and m, which is the same to the time complexity of simple linear model. Our method is being implemented into the widely used EPACTS (Efficient and Parallel Association Container Toolbox for Sequence data) software package, with many utilities and visualization supports for users. We believe that our extremely effective and useful software package to unravel the causal regulatory mechanism from DNA and mRNA sequencing data.

1916W
Using GenTrain and Z-Call to Identify Problematic SNPs in Rare Variant Genotype Calling. J. Romm, I. McMullen, M. Jewell, J. Zhang, E. Pugh, K.F. Doheny, Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality Next-Gen Sequencing (NGS), Genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. Many commercially available genotyping arrays (HumanExome, HumanOmniv2.5, HumanOmniv and all Illumina ‘plus exome’ arrays) contain SNPs in the database that are not present in the ensemble genomic data. Gene- and pathway-level methods can rely on annotation from standardized databases but pathway-level methods use a homogeneous pool. Conclusions: The interpretation of gene- and pathway-level methods is dependent upon their specific, inherent assumptions and the resources they rely on. Low type I error is important for gene-level methods, but pathway- and pathway-level methods may provide valuable insight into the ‘missing heritability’ of traits.

1918F
Detecting genetic heterogeneity in complex diseases with a weighted U statistic. O. Lu, C. Wei, R.C. Elston.1) Dept of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio.

For most complex diseases, a large proportion of the genetic variants remain undiscovered. While current research interests have shifted toward uncovering rare variants, gene-gene/environment interactions, and structural variations, the impact of genetic heterogeneity in human diseases has been largely overlooked. Converging evidence suggests that diseases with the same or similar clinical manifestations could have different underlying genetic etiologies. Most of the existing analytical approaches assume the disease under investigation has a homogeneous genetic cause and could, therefore, have low power if the disease undergoes heterogeneous biological pathways. In this paper, we propose a statistical approach, a heterogeneity weighted U (HWU) approach, for high-dimensional association analysis taking genetic heterogeneity into account. HWU can be applied to various types of traits (e.g., binary and continuous), and is designed for detecting heterogeneous genetic effects. Through simulations, we compared HWU with a non-heterogeneity weighted U (NHU) and the conventional generalized linear model (GLM). The results showed that HWU has substantial gain in power compared to NHU and GLM in the presence of genetic heterogeneity, while retaining a performance similar to that of NHU and GLM when the effects are homogeneous. Using HWU, we conducted a genome-wide analysis to study genetic heterogeneity in nicotine dependence. This genome-wide analysis of nearly one million SNPs from the Study of Addiction: Genetics and Environments (SAGE) took 5.8 hours, identifying heterogeneous effects of two new genes (i.e., CYP39A1 and VDAC3) on nicotine dependence.
Efficient and Accurate Population-Scale KIR Typing from SNP Chip Data.

KIR (killer immunoglobulin-like receptor) genes are of great interest in regard to resistance to viruses, autoimmune disease, reproductive conditions and cancer. Like HLA, they are highly polymorphic and some KIR molecules interact with HLA class I. Better understanding genetic associations of KIR with disease requires accurate typing of both HLA and KIR variants. KIR exhibit copy number variation (CNV) and haplotypes may comprise from 4 to 15 or more genes. This kind of diversity makes KIR typing expensive and time consuming. Thus to date KIR has been understudied.

SNP imputation for other SNPs is now routine, and performs extremely well, and SNP-based imputation for HLA is very good and becoming more widely used, but currently there are no imputation methods for KIR. To address this we have developed a high-throughput, accurate statistical imputation methodology for KIR typing using SNP variation data. We use a reference data set of approximately 600 haplotypes of European descent typed at 305 SNP loci covering 400kb in the vicinity of the KIR genes. These SNPs were typed on Immunochip. We fit a statistical model to these SNPs and performed extensive validation experiments, using cross-validation and in a separate group of individuals, that enables us to impute KIR alleles from SNP variation alone. Validation of KIR typing from SNP data has been applied to both population and pedigree data, to both single SNP and multi-SNP data as random and fixed effects, to gene expression and other omics data, to both estimation and prediction of genetic risk, to understanding the genetic architecture of complex traits, and to improve predictive ability. We also present a detailed comparison between pedigree and genomic estimates of heritability in the Amish to measure missing heritability.

PSEUDOMARKER 2.0: efficient computation of likelihoods using NOMAD.

We present results in both animal and human data. We show that the genetic architecture of complex traits can include significant non-additive variance such as dominance in milk traits estimated in 35,000 Holstein cows and in blood pressure in 3500 Amish subjects. We show that including sources of non-additive variance in estimating SNP effects can improve predictive ability. We also present a detailed comparison between pedigree and genomic estimates of heritability in the Amish to measure missing heritability.

MMAP: a comprehensive mixed model program for analysis of pedigree and population data.

The application of mixed models for genetic analysis has seen a sharp increase in the last several years. The power and flexibility of the mixed model has been applied to both population and pedigree data, to both single SNP and multi-SNP data as random and fixed effects, to gene expression and other omics data, to both estimation and prediction of genetic risk, to mapping both common and rare variants, and to understanding the genetic architecture of complex traits. Kernel machine methods, kriging, and non-parametric regression can be cast into the mixed model framework to take advantage of available machinery. The basic ingredients in the mixed model are covariance matrices that measure similarity between subjects based on genetic data. MMAP is an optimized and flexible mixed model analysis platform that incorporates a wide range of covariance structures such as additive, dominance, epistasis, maternal and imprinting using pedigree and/or genotype data and also allows users to define their own covariance structures. Likelihood calculations use multi-threaded optimized matrix libraries to handle multiple random effects. MMAP can import data from a variety of imputation programs to avoid data manipulation and IBS/IBD programs to build covariance structures.

MMAP uses a fast low-memory method to calculate additive and dominant genetic covariance structures using SNP data, which can be quite challenging for large data sets. For polygenic SNP analysis MMAP can store SNP-covariance products to reduce the complexity of the subsequent analyses with the same subjects to linear regression, independent of phenotype or covariates. We present results in both animal and human data. We show that the genetic architecture of complex traits can include significant non-additive variance such as dominance in milk traits estimated in 35,000 Holstein cows and in blood pressure in 3500 Amish subjects. We show that including sources of non-additive variance in estimating SNP effects can improve predictive ability.
1922W
SALAD: A software suite for admixture linkage analysis and discovery.
R. Johnson1,2, G. Nelson1, C. Winkler3. 1) BSP CCR Genetics Core, Freder-
ick National Laboratory, SAIC Frederick, Inc, Frederick, MD; 2) Chaire Bioinf-
formatique, Conservatoire National des Arts et Métiers, Paris, France; 3) Basic Research Laboratory, Frederick National Laboratory, SAIC Frederick, Inc, Frederick, MD.

The importance of accounting for population structure in genetic studies and the power of admixed populations has long been recognized. Technolog-
cal advances over the last decade have made it possible to sample and analyze these populations for regions of admixture linkage disequilibrium associated with clinical outcomes. Admixture linkage analysis remains a powerful, important tool in genetic studies, and many good software pack-
egages exist to aid the researcher. Multiple software packages requiring dispar-
ate data formats are required to fully analyze and prepare all but the simplest admixture linkage models. The Suite for Admixture Linkage Analysis and Discover (SALAD) was developed to bring together tools for admixture linkage studies from study design to manuscript preparation. The backbone of SALAD, local and global ancestry inference, is done using a Hidden Markov Model similar to other admixture linkage software. These ancestry estimates can be further analyzed in a flexible modeling framework, allowing the user to model clinical outcomes as a function of global and local ancestry in addition to relevant covariates, using the appropriate generalized linear model. The case-only locus-genome statistic implemented in ANCES-
TRYMAP (Patterson et. al. 2004) is also included, as well as basic locus-
genome statistics for other data types (e.g. continuous or survival). This can be generalized to compare any two user defined models based on Bayes factors. Also included in SALAD are tools for ancestry informative marker panel design, admixture linkage disequilibrium analysis of common SNPs (GWAS data analysis and various publication quality graphics functions. Power and accuracy of SALAD is tested using simulated data, and relevant comparisons with similar software packages are made.

1923T
Local Ancestry Inference for Whole Genome Sequence Data.
B. Maples1,2, S. Gravel3, E. Kenny4,5,6,7, C. Bustamante1. 1) Genetics, Stanford University, Stanford, CA; 2) Biomedical Informatics, Stanford University, Stanford, CA; 3) Human Genetics, McGill University, Montreal H3A 1B1, Canada; 4) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Statistical Genetics, Icahn School of Medicine at Mount Sinai, New York, NY; 6) Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY.

Local ancestry inference is an important step in the genetic analysis of fully sequenced human genomes. Current methods can only detect continent-
Ial-level ancestry (i.e., European vs. African vs. Asian) accurately even when using millions of markers. Here, we present RFMix, a powerful discriminative modeling approach that is faster (~30X) and more accurate than existing methods. We accomplish this by using a conditional random field (CRF) parameterized by random forests trained on reference panels. RFMix is capable of learning from the admixed samples themselves to boost perform-
ance and autocorrect phasing errors. RFMix shows high sensitivity and specificity in simulated Hispanic/Latinos, African Americans, and admixed Europeans, Africans, and Asians. Finally, we demonstrate that African American in HapMap contain modest (but non-zero) levels of Native American ancestry (~0.4%).

1924F
Estimating Sample Size and Power for Nested Cohort Studies. R. Hoff-
mann. QHS. Pediatrics, Medical College of Wisconsin, Milwaukee, WI.

Objective: Because of the rarity of certain diseases, the high cost of gene sequencing and the high cost of obtaining gene expression data the sample size and power of a nested cohort study needs to be obtained when all the cases, but only some of the non-diseased are sampled.

Methods: A nested cohort study may sample all the cases in each cohort, but only some of the controls from each cohort. Obtaining estimates of Relative Risk (RR) or the expression ratio involves choosing the correct sample from the non-diseased from both cohorts. The statistical power was determined separately for matched, stratified (using a stratified random sample) and unmatched (using a simple random sample) non-diseased. The power was determined using simulated data with both rare variant and common variants making up a complex disease structure. Rare variants are defined as alleles that have <0.5 percent, <1 percent or less than 5 percent prevalence. The disease structure is patterned on either (1) a disease structure that has a few well-defined candidate genes, as well as a disease structure that potentially involves as many as 5000 potential locations. Stratifi-
cation ranged from no stratification to stratification that is related to the disease to stratification that is unrelated to the disease.

Results: A sample of matched non-diseased to the diseased is found to have the least power and the highest probability of a false discovery. A stratified random sample had high power and a substantially lower probability of false discovery. An unmatched sample was found to have substantially higher power, but a higher probability of false discoveries if substantial stratification is present and not related to the disease system. Continuous expression data had higher power, but otherwise showed the same pattern with respect to sample selection of the non-diseased.

1925W
Impact of measurement error on testing genetic association with quantita-
tive traits.
J. Liao1,2, X. Li3,4, T-Y. Wong4,5,6,7, J.J. Wang8, C.C. Khor3, E.S. Tai4,5, T. Aung1,2, Y.Y. Teo8, C.Y. Cheng1,2,4,8. 1) Department of Ophthal-
mology, National University of Singapore and National University Health System, Singapore, Singapore; 2) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; 3) Department of Statistics and Applied Probability, National University of Singapore, Singa-
pore, Singapore; 4) Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore; 5) Centre for Vision Research, University of Sydney, Sydney, Austra-
ilia; 6) Division of Human Genetics, Genome Institute of Singapore, Singa-
pore, Singapore; 7) Department of Medicine, National University of Singa-
pore and National University Health System, Singapore, Singapore; 8) Duke-
NUS Graduate Medical School, Singapore, Singapore.

Background: Ignoring measurement error of a phenotypic trait will lead to underestimation of the standard error and hence reduce the power to detect genetic associations. We examined the impact of sample size, allele frequency and effect size in the presence of measurement error for quantita-
tive traits.

Methods: The statistical power to detect genetic association by comparing phenotype mean and variability was investigated analytically. The non-cen-	rality parameter for a non-central F distribution was derived and verified using simulation study. We obtained equivalent formulas for the cost of measurement error of a standard normal distributed trait required a one-fold increase in sample size for comparison of means, and a three-fold increase in sample size for comparison of variances. GWAS results revealed almost no overlap in the significant SNPs (p <10−6) for the two cataract grading scales. Genetic variant rs1458038 associated with blood pressure was replicated in aver-
aged blood pressure measurements but not in single blood pressure mea-
surement due to larger measurement error.

Conclusions: We have developed a framework for researchers to quantify power or to calculate sample size in the presence of measurement error, which will be applicable to studies of phenotypes in which the measurement is highly variable.
1926T Slicing the Genome: A New Approach to Association in Complex, Longitudinal Diseases. A. Musolf1, D. Londono1, A.Q. Nato, Jr.2, P. Vuister2, C.A. Wise2, S.L. J. Finch3, M. Bochud4, T.C. Matise3, D. Gordon1. 1) Department of Genetics, Rutgers University, Piscataway, NJ, USA; 2) Division of Medical Genetics, University of Washington, Seattle, WA, USA; 3) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 4) Sealy Center forMusculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, TX, USA; 5) Department of Orthopedic Surgery, Texas Scottish Rite Hospital for Children, TX, USA; 6) Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA; 7) MD Studies, P.R.C. Research Center, Marshfield, WI, USA; 8) Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA; 9) Unit for Prevention and Control of Cardiovascular Disease, Section of Non Communicable Diseases, Ministry of Health and Social Services, Seychelles.

We previously published a method that tests for association between a longitudinal phenotype and genetic variants. The method uses growth mixture models (GMM) to determine longitudinal trajectory curves. The Bayesian posterior probability (BPP) of belonging to a specific curve, an outcome variable from the GMM, is used as a quantitative phenotype in association analyses. Though the method proves to be powerful for a single causal variant under multiple inheritance scenarios, power significantly decreases when more than one causal variant is considered. Here, we present a new method designed to detect multiple causal SNPs associated with longitudinal phenotypes in both family and population studies. The method also allows for the incorporation of covariates. This novel method retains several ideas from our first method, however instead of performing individual association tests with a single phenotype, we test for association with blocks of 50 SNPs (which we term a "mega-locus") and obtain a significance value on each mega-locus. This is accomplished via the SumStat method, developed by Jurg Ott and colleagues. As SumStat works for population studies only, we use a modified procedure (TDT-HET) to test for family-based association. We consider various scenarios in our simulations, including four causal variants located within a single mega-locus and eight causal variants spread between two mega-loci on different chromosomes. We also introduce environmental covariates. Our data set is highly stratified to ensure representativeness in the presence of population stratification. P-values for each mega-locus on each data set are computed. To adjust for multiple testing, the final p-values are combined via Fisher's method (per mega-locus) and by the false discovery rate (FDR). We report that our simulations: 1) appear to maintain the proper type I error and 2) have greater than empirical 75% power for most simulations. These results suggest that our method can detect multiple causal SNPs located in multiple regions across the genome. We believe that this method will be useful to researchers who are studying complex diseases with longitudinal phenotypes. It allows for potentially high power for association of causal loci with disease progression phenotypes for both population and family studies, even in the presence of confounding elements such as population stratification and environmental variables.

1928W Optimal Strategies for Identifying Disease Associated Singletons. S. Rashkin, G. Jun, G. Abebaxis. Center Statistical Genetics, Univ Michigan, Ann Arbor, MI, USA; 1) Department of Biostatistics, University of Washington, Seattle, WA, USA; 2) Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada; 3) Department of Statistics and Actuarial Science, University of Waterloo, Waterloo ON, Canada.

Rare functional variants are hypothesized to explain much of the heritability of common, complex diseases. These variants are generally missing from genotyping array based studies and understanding their role in disease requires sequencing. Singletons, the vast majority of these variants, can only be detected with high power by deep sequencing, which remains expensive. At lower depth, costs can be greatly reduced, enabling sampling sizes to increase, but many singleton variants will be missed. Here, we examine the balance between power to detect rare variants at different sequencing depths and sample sizes, so as to maximize power and minimize cost for studies that explore the role of singleton variants in human disease.

We used a simulated multismaple caller to calculate power of singleton discovery based on different read depths (2-50x), sample sizes (20-5000), and error rates (0.001-0.01). While all three factors influence power, read depth appears to have the largest effect. At 10x coverage, power to detect a singleton variant is greater than 0.75 for all combinations of sample sizes and error rates examined. At 20x coverage, power exceeds 0.95, regardless of sample size and error rate. We validated our simulation by down-sampling deeply sequenced exome samples and assessing our ability to detect previously called singletons.

While we are interested in our ability to detect singletons, it is also important to study how this affects association results. We approximated the power of an association test using analytical computations that considered sample size, population frequency of singletons, relative risk for carriers, prevalence of disease, and read depth. Analysis of association study power allows for greater insight into the ideal sample size and read depth to use in studying singletons under a variety of conditions. Changes in sample size, relative size, or frequency of singletons have a large effect on power: increasing frequency of singletons or relative risk allows for smaller sample sizes to attain high power. For a fixed sample size, at lower depth (<10X), increasing coverage increases power; as depth increases (typically beyond >15-20X), power remains relatively constant. Thus, our results suggest that studies of the contribution of rare variants to human disease will be optimally powered at coverage data sets – but that coverage should only be increased when such an increase does not require a decrease in sample size.

1929T Response Dependent Sampling Designs And Analysis In Studies With Rare Variants. L. Sun1,2, A. Danksch2, J.F. Lawless3,1. 1) Division of Biostatistics, Dalra School of Public Health, University of Toronto, Toronto, ON, Canada; 2) Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada; 3) Department of Statistics and Actuarial Science, University of Waterloo, Waterloo ON, Canada.

Rare variants play an important role in studies of complex human diseases and traits, and next generation sequencing technology provides rich data for analysis. This recent focus on rare variants has produced a significant number of sequencing studies as well as a large number of association tests. However, theoretical and empirical results have shown that in order to achieve high power large sample sizes are needed. Compared to random sampling, it has been shown that response-selective sampling designs can greatly improve the power of association studies with common variants and recently, rare variants. We extend the range of tests for multiple rare variants and show how adjustments to existing tests can increase power. We also investigate sampling strategies where random samples are taken from the tails of the response distribution and there may also be stratification on covariates. Through theoretical calculations and extensive simulations, we investigate what designs and testing approaches are efficient and robust across various genetic models.

1927F Use of Electronic Medical Records to Measure Phenotypic Heritability. S. Hebbingt1, Z. Ye1, M. He2,3, J. Mayer4, S. Schrodi1, M. Brilliant1. 1) Center for Human Genetics, Marshfield Clinic, Marshfield, WI; 2) Bioinformatics Research Center, Marshfield Clinic, Marshfield, WI.

A priority of the NIH is to understand the genetic etiology of common and rare diseases so that genetics may be used in ‘personalized medicine.’ All diseases are the result of a combination of environmental and/or genetic factors. Conducting genetic/genomic studies is difficult without evidence of a strong genetic component by heritability measurements. Heritability can be measured in a variety of family structures including twins and extended families. A particular challenge in conducting heritability estimates is identifying appropriate families with available phenotypic information. To address this challenge, and to demonstrate proof-of-principle that electronic medical records (EMRs) can be used to efficiently characterize disease heterogeneity, we estimated the heritability of muscular dystrophy (MD) in a Marshfield Clinic patient cohort (Personalized Medicine Research Project, PMRP) with available family structure data (4,475 families) all linked to patient EMRs. MD was chosen as a positive control because of its well characterized genetic etiology and has a well-defined ICDO code (359.1) easily extractable from the EMR. In this population, 6 out of 18 MD cases were familial in nature. The broad sense heritability estimated for MD using the EMR was 0.63 (SE 0.16), and closely matched expected estimates. To further follow-up on the use of EMRs to characterize the potential genetic contribution of human diseases, we established a Marshfield Clinic twin cohort (15,802 twin pairs) and again ascertained MD diagnosis using the EMR. In the twin cohort, there was an enrichment for disease concordance in twins (4 affected pairs out of 17 individuals, p=1.59E-8). These results, both from extended and twin cohorts, demonstrate proof-of-principle that the EMR may be a powerful and efficient tool when doing studies of heritability even in the absence of existing genetic/genomic data.
1930F Quantitative-trait-dependent sampling designs for genetic association analysis of a rare variant score. Y.E. Yilmaz1,2,*, Y.F. Lawless1,2,*, S.B. Bull1,2, 1) Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3) Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada; 4) Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John’s, Newfoundland, Canada.

For rare variant analysis related to a quantitative trait (QT), selection of individuals for sequencing according to their QT value can improve cost-efficiency. We consider QT-dependent two-phase sampling designs for regression analysis of a QT on a rare variant score obtained by aggregating rare genetic variants in a specified functional unit. In the first phase, we obtain trait values for all individuals in a cohort; in the second phase, we obtain genetic sequence data for a subset of individuals selected according to their trait values. Under stratified sampling, individuals in strata defined by high or low trait values can be selected with a higher fraction than those with an intermediate trait value. Under extreme-trait sampling, only individuals with high or low trait values are selected. In such trait-dependent sampling designs, inference based on standard linear regression methods that treat the QT as the dependent variable, ignoring the selection, may be misleading. We review the key features of two well-known methods for analysis under two-phase designs: likelihood-based inference for response-biased samples and inverse probability weighting (IPW) based on estimating equations, and describe analytic and simulation-based approaches to compare various sampling designs according to the number and size of strata, and the allocation of the phase 2 sample to the specified strata, as well as the distribution of the rare variant score and associated effect size. The relative efficiency of alternative designs depends on the method of analysis. Under a stratified sampling design, and assuming a linear relationship between the trait and rare variant score with normally distributed error term, likelihood inference favours extreme-trait-selection whereas IPW analysis requires non-zero sampling fractions in all strata and can be sensitive to sparsely sampled strata. Based on evaluation of type I error and power of a likelihood ratio test of rare variant association by simulation, we recommend likelihood-based inference for extreme-trait sampling and more general stratified sampling, with the choice between them based on considerations of robustness as well as relative efficiency.

1931W Caveats of extreme sampling strategies for resequencing studies. H. Qin1, W. Ouwyang1, S. Cao1,2, T. Yang1, Y.-P. Wang1,2, H.-W. Deng1. 1) Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, New Orleans, LA 70112, USA; 2) Department of Biomedical Engineering, Tulane University, 534 Lindy Boggs Building, New Orleans, LA 70118, USA.

Selective sampling strategies are crucial to enrich causal rare genetic variants for cost-effective next-generation resequencing studies. One widely employed selective sampling strategy is extreme phenotype sampling - sampling individuals with extreme phenotypes. It seems to be a common sense that extreme phenotype sampling can enrich the presence of causal rare genetic variants compared to random sampling. An intuitive alternative of extreme phenotype sampling is extreme residual sampling - sampling individuals with extreme residuals after adjusting for covariates (e.g., identified genetic variants, population structure surrogates, and environmental factors) from phenotypes. In this paper, we mathematically modeled the causal architecture among phenotype, target genetic variants, and covariates. Under the causal architecture, we analytically investigated the probabilities of the three sampling strategies to detect target genetic variants. We proved the following results: Extreme phenotype sampling is inferior to random sampling when covariates are negatively correlated with target genetic variant. Extreme residual sampling is superior to random sampling; but it is inferior to extreme phenotype sampling if covariates are positively correlated to target genetic variant and target genetic effect is mild. Typical examples are incorporated for illustrations. In conclusion, no sampling strategy can uniformly dominate all the others under all circumstances. For a resequencing study, the ‘optimal’ sampling strategy depends on the underlying causal architecture among phenotype, target genetic variants, and covariates.

1932T Identifying potential cancer vaccine targets with high-throughput sequencing. E. Aronesty1,2,*, K. Robasky1, W.D. Jones1,2, 1) Bioinformatics, Johns Hopkins University, Washington, DC; 2) Expression Analysis, a Quintiles Company, Durham, NC; 3) Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC.

Since the HPV vaccine was introduced to prevent cervical cancer, other vaccine targets remain elusive. Here, we use RNA-Seq to detect viral homologs in tumor sample expression data obtained from TCGA. We compare with normal tissue expression obtained from GTEX data sets, and to ‘paired-normal’ from blood samples where available. We have inspected RNA sequence of for breast, liver, cervical, ovarian, prostate, colon, lung, and brain cancers, and found that the overall viral presence is higher in cancer tissue compared to unaffected patient tissue. We present a method for producing a virome, removing low-complexity reads, counting viral presence by genus, as well as two methods for comparing counts between patient tissue and GTEX reference tissue. Finally, we assemble contigs for statistically significant viruses, and annotate results in a manner sufficient for vaccine target identification.

1933F A Dynamic Model for Classification of Gene Regulation with RNA-seq Data. L. Li, M. Xiong. University of Texas School of Public Health.

Characterizing gene regulation and capturing its feature will provide valuable information for understanding biological processes. Variation in gene expression underlies many biological processes and holds a key to unraveling the mechanism of gene regulation. However, the gene expression measured by microarray provides limited information to reveal the features of gene regulation. The rapidly developed NGS technologies have been becoming the platform of choice for gene expression profiling. RNA-seq for expression profiling offers comprehensive picture of transcriptome and has made a number of significant qualitative and quantitative improvements on gene expression analysis and provides multiple layers of resolutions and transcriptome complexity: the expression at exon, SNP, and positional level; splicing; post-transcriptional RNA editing across the entire gene; isoform and allele-specific expression. To unravel the features of gene transcription, we propose to use differential equation to model the observed number of reads across the gene. We view the number of reads or expression level at each position as a function of the genomic position and view the transcription process as a dynamic process of transcription along the genome. Instead of taking derivative of expression level with respective the time, we calculate derivative of the expression level with respect to genomic position. We use a second order differential equation to model the dynamics of transcription process along the genome. Iterative principal differential analysis is used to estimate the coefficients in ODE by specifying the ODE as the data-driven penalty in the B-spline smoothing. We iterate between curve smoothing and ODE estimation until convergence occurs. The proposed methods are applied to ovarian cancer RNA-seq data with 412 tumor samples from TCGA dataset. We study the stability and transient response of transcriptional process for each gene using its fitted differential equation. The dynamic features of the gene transcription can reveal various alternative splicing, alternative start and end of transcription, and isoforms. Based on the dynamic features of gene transcriptional process we use the coefficients of ODE to classify the gene transcription into four categories. Our results will provide valuable information for understanding the mechanism of gene regulation and unraveling disease process. These results may also open a new way to find drug target and disease treatments.
Telomere Length in Circulating Leukocytes Is Associated with Lung Function and Disease. E. Albrecht1, E. Sillanpää1, S. Karrasch1, A. Couto Alves2, V. Corda1,2,4, H. Hovatta1,2,4, T. D. Buxton3, S. Hägg3,10,11, M. Mangino12, G. Willemsen13, K.H. Pietiläinen14,15, C.P. Nelson16,17, L. Broe16,17, M.A.R. Ferreira18, I. Surakka15,19, C. Gieger1, N.G. Martin16, N.L. Pedersen10, D.I. Broomsma1, T.D. Spector1, C.M. van Duijn15,17, J. Kaprio15,16, N.J. Samani14, M.R. Jarvelin1,4,21,22,23,24, H. Schulze25 1) Institute of Genetic Epidemiology, Helmholtz Zentrum Münchern, Neuherberg, Germany; 2) Gerontology Research Center and Department of Health Sciences, University of Jyväskylä, Finland; 3) Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich, Germany; 4) Dept of Epidemiology and Biostatistics, MRC-HPA (Health Protection Agency) Centre for Environment and Health, School of Public Health, Faculty of Medicine, Imperial College London, UK; 5) Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; 6) Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester, UK; 7) Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland; 8) Department of Biosciences, Viikki Biocentre, University of Helsinki, Finland; 9) Department of Medicine, Imperial College London, UK; 10) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 11) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 12) Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK; 13) Dept of Biological Psychology, VU University Amsterdam, Amsterdam, the Netherlands; 14) Obesity Research Unit, Department of Medicine, Division of Endocrinology, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; 15) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Finland; 16) Netherlands Consortium for Healthy Aging, Leiden University Medical Center, Leiden, the Netherlands; 17) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 18) Centre for Epidemiology and conditions of Medical Research, Brisbane, Queensland, Australia; 19) Public Health Genomics Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland; 20) University of Helsinki, Hjelt Institute, Dept of Public Health, University of Helsinki, Helsinki, Finland; 21) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Finland; 22) Unit of General Practice, University Hospital Oulu, Finland; 23) Biocenter Oulu, University of Oulu, Finland; 24) Department of Lifecourse and Services, National Institute for Health and Welfare, Oulu, Finland; 25) Institute of Biostatistics, University of Oulu, Finland; 26) German Research Center for Environmental Health, Neuherberg, Germany.

Clinical and experimental studies suggest the involvement of premature aging in COPD. Using an epidemiological approach we studied whether accelerated aging indicated by reduced telomere length, as a marker of biological age, is associated with COPD and asthma, and whether intrinsic age-related processes contribute to the inter-individual variability of lung function. Our meta-analysis of 14 studies included >1,000 COPD cases with >15,000 controls, >2,500 asthma cases with >28,000 controls, and the spirometric indices forced expiratory volume in one second (FEV1), forced vital capacity (FVC), and their ratio FEV1/FVC in >13,000 individuals. Associations were tested by linear regression, adjusting for age, sex, and smoking status. We observed negative associations between telomere length and COPD as well as asthma, with stronger effects in women compared to men. The analysis of spirometric indices showed positive associations between telomere length and all investigated spirometric measures. The associations were weaker in apparently healthy subjects compared to COPD or asthma patients. Our results indicate that lung function may reflect biological aging primarily due to intrinsic processes which are likely to be aggravated in lung diseases. Shortened telomeres in COPD and asthma suggest that induced aging is involved in the pathogenesis of these diseases.
1937W


1) Public Health Genetics, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Carolina Center for Genome Research, University of North Carolina, Chapel Hill, NC; 3) Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, NC; 4) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 5) Department of Genetics, Rutgers University, Piscataway, NJ; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 7) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 8) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN.

Type 2 Diabetes Mellitus (T2DM) has an increasing burden of morbidity and mortality worldwide, particularly in populations of non-European descent. Although Genome-Wide Association Studies (GWAS) have established numerous loci associated with diabetes-related glycemic traits, the subsequent identification of causal variants remains a considerable challenge. Leverage differences in local linkage disequilibrium (LD) structure between genetically diverse populations is a powerful method to narrow candidate regions. To explore the generalizability of susceptibility loci for fasting glucose (FG) and fasting insulin (FI) we fine-mapped 17 previously identified loci by investigating a total of 12,227 variants genotyped on the MetaboChip in individuals without diabetes from the PAGE study. A previous analysis demonstrated that exploring European GWAS loci in African Americans is a powerful approach to identify likely functional variants. We developed signals for FG and FI within PAGE. In this analysis we expand upon previous work to investigate the generalizability of European GWAS loci and search for novel SNP effects for FG and FI across multiple ethnicities in a total of 21,000 Asians, 8140 African, 3500 Hawaiian and 11,785 African American individuals. Association between SNPs and trait was evaluated using linear regression for FG and logistic regression for ln(FI) across an additive genetic model (adjusted for age, sex, BMI, diabetes status, and principal components). Each ethnic study was analyzed separately. The analysis was conducted on results across studies within each ethnicity and with all ethnicities combined. In our preliminary analysis we attempted to fine-map 17 European GWAS loci by genotyping an average of 720 SNPs per region in 3,215 Hispanic women from the Women’s Health Initiative. Using a Bonferroni adjusted threshold, we detected one significant SNP (rs116132956, p=1.29 mmol/L, p=6.12x10^-7) in the DGKB region previously identified in AA and EA. Although the direction of effect was the same across populations, the most extreme p-values were at three statistically independent SNPs for AA, HA, and EA suggesting fine mapping of functional variants at the DGKB locus. A deeper understanding of the genetic influences across populations on glucose and insulin levels will provide insight into the underlying biology of T2DM that could have important implications for the development of therapeutic targets and prevention strategies.

1939F

Derivation of a genome-wide significance threshold for African populations. M.D. Fortune1,2, I. Tachmazidou1, E. Zeggini1.

1) Welcome Trust Sanger Institute, Hinxton, UK; 2) Department of Applied Mathematics and Theoretical Physics, Cambridge, UK.

Genome-wide association studies examine common variation across the genome for association with complex traits of interest. Significance is declared at the widely-accepted threshold of p<5.0E-08. This has been derived from the total number of effective common variant (minor allele frequency (MAF) >0.05) tests in European populations and has been based on HapMap data. As the GWAS field is shifting to the study of more structured and heterogeneous populations, for example from a genetic descent, a new statistical significance level has to be defined. Lower levels of linkage disequilibrium between common variants may necessitate a more stringent threshold. In addition, the availability of sequence data further empowers the assessment of the effective number of independent tests, as common variation has been comprehensively assayed. Many methods exist which exploit the correlation structure, either haplotypic or genotypic, between the variants. We have implemented several of these on two African datasets, Luhya in Webuye, Kenya (LWK) and Yoruba in Ibadan, Nigeria (YRI). From the 1000 Genomes Project (sequence data, phase I integrated public data release), in order to estimate the effective number of tests for common genetic variation (MAF over 1 or 5%) in African populations. For comparison we also used the Utah residents (CEPH) with Northern and Western European ancestry GUE dataset from the same source. Using the haplotype correlation coefficients, as proposed by Moskvina and Schmidt resulted in an estimate of 3.0E-8 for the European dataset, and 1.15E-8 for the African datasets at MAF over 5%. The same algorithm at MAF over 1% gave an estimate of 3.5E-8 for LWK and 3.5E-8 for YRI. This reflects the greater genetic variation in present day sub-Saharan African populations.

1940W

Mediating genetic effects using twin data. A. Ulgen1, W. Li2, J. Hjelmborg2.

1) Faculty of Medicine, Eastern Mediterranean University, Famagusta, TRNC, Mersin-10-Turkey; 2) Robert S Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore Long Island Jewish Health System, Manhasset, New York, United States of America; 3) Department of Biostatistics, Institute of Public Health, University of Southern Denmark, Odense, Denmark.

We apply a genetic modeling via simulation to twin data for obesity and cancer related measurements. For obesity, we use both BMI and other quantities derived from BMI to measure the weight growth. More specifically, the log(BMI) at baseline is regressed over other factors: ln(BMI_i_j) = beta_1 + alpha_1 i + beta_2 + alpha_2 j + epsilon_i_j, for individual i and timepoint j (time since the baseline, in years). The alpha_1 is then the log-weight growth rate (Hjelmborg et al. Obesity, 16(4), 2008). For cancer data, we use phenotypes available in the registry. We follow the genetic modeling of twin data proposed by (Dite et al. and Stone et al; Cancer Epidemiol Biomarkers Prev; 17(12), 2008 and 17(12), 2012, respectively.) In this modeling, the phenotype of a twin in a twin pair is regressed over both twins’ co-variates. If the two twins in a twin pair are labeled as 1 and 2, Y denotes phenotype and X co-variates, then E(Y)=alpha + beta_1 X_1 + beta_2 X_2. It was shown that by varying a co-variate, the expected value of the phenotype measure would change. In our analysis, we assume a bivariate normal distribution for both (Y1,Y2) and (X1,X2). We treat phenotype measurements such as BMI, growth rate, at the baseline time as X, and that at the later time as Y. This approach would incorporate measurements at two points along a time course, thus enhance the power to detect the genetic component. We also introduce a random effects model for the stratification effects.

1938T


1) Department of Biology, Indiana University, Bloomington, IN; 2) Department of Biostatistics, Institute of Public Health, University of Southern Denmark, Odense, Denmark; 3) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN.

We present a method to estimate genome-wide genetic correlations from two traits, is a problem of high interest (Lee et al. 2012 Nat Genet). Here, we demonstrate that exploring European GWAS loci in African Americans is a powerful approach to identify likely functional variants. We developed signals for FG and FI across an additive genetic model (adjusted for age, sex, BMI, diabetes status, and principal components). Each ethnic study was analyzed separately. The analysis was conducted on results across studies within each ethnicity and with all ethnicities combined. In our preliminary analysis we attempted to fine-map 17 European GWAS loci by genotyping an average of 720 SNPs per region in 3,215 Hispanic women from the Women’s Health Initiative. Using a Bonferroni adjusted threshold, we detected one significant SNP (rs116132956, p=1.29 mmol/L, p=6.12x10^-7) in the DGKB region previously identified in AA and EA. Although the direction of effect was the same across populations, the most extreme p-values were at three statistically independent SNPs for AA, HA, and EA suggesting fine mapping of functional variants at the DGKB locus. A deeper understanding of the genetic influences across populations on glucose and insulin levels will provide insight into the underlying biology of T2DM that could have important implications for the development of therapeutic targets and prevention strategies.

1939T


1) Department of Biology, Indiana University, Bloomington, IN; 2) Department of Biostatistics, Institute of Public Health, University of Southern Denmark, Odense, Denmark.

We present a method to estimate genome-wide genetic correlations from two traits, is a problem of high interest (Lee et al. 2012 Nat Genet). Here, we demonstrate that exploring European GWAS loci in African Americans is a powerful approach to identify likely functional variants. We developed signals for FG and FI across an additive genetic model (adjusted for age, sex, BMI, diabetes status, and principal components). Each ethnic study was analyzed separately. The analysis was conducted on results across studies within each ethnicity and with all ethnicities combined. In our preliminary analysis we attempted to fine-map 17 European GWAS loci by genotyping an average of 720 SNPs per region in 3,215 Hispanic women from the Women’s Health Initiative. Using a Bonferroni adjusted threshold, we detected one significant SNP (rs116132956, p=1.29 mmol/L, p=6.12x10^-7) in the DGKB region previously identified in AA and EA. Although the direction of effect was the same across populations, the most extreme p-values were at three statistically independent SNPs for AA, HA, and EA suggesting fine mapping of functional variants at the DGKB locus. A deeper understanding of the genetic influences across populations on glucose and insulin levels will provide insight into the underlying biology of T2DM that could have important implications for the development of therapeutic targets and prevention strategies.
Detecting recent coevolution through ancestry association on different chromosomes in African-Americans. H. Wang¹, Y. Choi², X. Wang³, B. Tayo⁴, U. Brockel⁵, C. Hanis⁶, S. Karda⁷, S. Redline⁸, R. Cooper⁴, N. Risch⁹, H. Tang¹, X. Zhu¹. ¹) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; ²) Department of Genetics, Stanford University, Stanford, CA; ³) Department of Biostatistics, Harvard, Boston, MA; ⁴) Department of Public Health Science, Loyola University Medical Center, Maywood, IL; ⁵) Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI; ⁶) Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Science Center at Houston, Houston, TX; ⁷) Department of Epidemiology, University of Michigan, Ann Arbor, MI; ⁸) Department of Medicine, Harvard Medical School, Boston, MA, USA; ⁹) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA.

Genetic coevolution could be maintained through compensatory mutations over evolutionary history. Recent studies have found evidence of coevolution between physically unlinked genes and loci. We hypothesized that recent allelic coevolution could be detected by examining simultaneous or complementary transmission of ancestry alleles in different genomic regions. This hypothesis can be tested in recently admixed populations such as African Americans. We analyzed the correlations of local ancestry across the genome in 20,097 African Americans sampled from three different study cohorts: CARe, WHI, and FBPP. The ancestral proportions were estimated using HAPMIX and SABER+. Inverse-variance weighted meta-analysis approach was used to combine results from the three study cohorts. Our analysis revealed 79 pairs of genetic regions on different chromosomes which are significantly correlated (p-value < 10⁻⁹ in meta-analysis), suggesting possible evidence of coevolution. Several of these regions contain genes previously demonstrate to be under selection such as SLC30A9 and ASPM.

Investigating bias due to population stratification in pharmacogenetic studies. X. Chen, C. Molony, C. Zhang, H. Zhou. Informatics and Analysis, Merck, Boston, MA.

It is well known that bias will be introduced in population-based association studies in the presence of population stratification (PS). Failing to control for PS might lead to identification of spurious association signals from ancestry-informative markers, rather than the causal disease loci. However, the impact and magnitude of this confounding by ethnicity remain unclear in the context of pharmacogenetic studies (PGx), where genome-wide associations studies (GWAS) are applied retrospectively on a subset of population from Phase II/III randomized clinical trials to discover genetic biomarkers associated with drug responses. Through using a simulation study design, we systematically investigated the type I error rates and bias in the estimates caused by PS in PGx. Various population substructures and treatment response models were generated to evaluate the bias in main genetic effects and the interactions between gene and treatment effects. Simulation results indicated that it is crucial to perform association analyses within each genetically homogeneous population even in well conducted randomized clinical trials. Otherwise false positive results due to differences in drug response rates across ethnicities were likely to be observed. With respect to the Phase II clinical trials conducted in a mixed population (e.g. African American), the bias due to PS appeared limited when differences of drug response rates among ethnicities are small to moderate. We also demonstrated that applying standard statistical approaches of adjusting ancestral histories can effectively provide unbiased estimations of main genetic effects and genetic by treatment interaction effects in PGx for identifying predictive biomarkers of drug response, while maintaining valid false-positive rates.
Validation of an ancestry estimation method using a comparison of FRAPPE and STRUCTURE. J.J. Bryan1, K. Tang1, R. Kittles2, C.L. Mounts1. 1) Sorenson Genomics, 2495 South West Temple Salt Lake City, UT. 84115; 2) University of Chicago 5841 South Maryland Avenue AMB W601 MC6091 Chicago, IL 60637.

An analytical test method for estimating genetic ancestry has been developed wherein 190 Ancestral Informative markers (AIMS) are genotyped and compared to reference populations. This study was performed to ascertain the test's ability to accurately estimate admixture. Ancestry estimation was performed for 44 unrelated, self-declared 'African American' (ASW) samples from the HAPMAP3 dataset using a genome-wide dense marker set (~860k SNPs) to establish the 'true' genetic affinities to 5 reference populations (European, East Asian, India Subcontinent, Indigenous America, and Sub-Saharan Africa). The same set of 44 samples was analyzed again, using the 190 AIMS in our test. For the 190 AIMS, two different statistical programs were independently employed, FRAPPE and STRUCTURE, both recognized for their validity in this type of comparative analysis. Each sample analysis was performed 50x, with a permutation procedure to account for the variability in the reference sample panels. The genetic affinities were thus estimated with well defined confidence intervals. The ancestry estimation results with the 190 AIMS were then compared to that of the high density ~860k set and then to each other. The results showed that our ancestry tests with either method corresponded very well to the data from the ~860k estimations; and strong consistency was also observed between the FRAPPE and STRUCTURE programs. More than 90% of the samples having a comparative r-squared value greater than 0.92 between the two programs. The overall difference between results obtained using FRAPPE and STRUCTURE were minimal, with an average difference between the corresponding ancestry percentages less than 0.2%. These data were further validated using an additional 144 self-declared 'African American' samples, with near equivalent results.

Geographic Population Structure (GPS) of worldwide human populations infers biogeographical origin down to home village. E. Elhaik1, T. Talatrina2, D. Chebotarev2, S. Piras3, C.M. Calo5, R.I. Montis5, M. Aztori5, M. Marin1, S. Tofanelli5, P. Francalacci5, L. Pagani6, C. Tyler-Smith8, Y. Xue6, G. Cucca4, T.G Schurr4, J.B. Gaieski7, C. Melendez2, M.G Villar8, R. Gomez11, R. Fujita11, F.R. Santos12, D. Comas13, O. Balanovsky4, W. E. Smith20, T. G. Schurr18, M. Hammer10, L. Matiasson-Smith20, S.R. Wells4, 1) Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205; 2) Glamazon Computational Biology Research Group, University of Glamorgan, Wales, CF371 PP; United Kingdom; 3) Laboratory of Applied Pharmacogenetics and Genomics, Children's Hospital Los Angeles, University of Southern California, 4650 Sunset Blvd, Los Angeles, CA 90027; 4) Department of Sciences of Life and Environment, University of Cagliari, Monserrato, SS 554, 09042, Italy; 5) Research Laboratories, bcs Biotech S.r.l., Viale Monastir 112, 09122 Cagliari, Italy; 6) Department of Biology, University of Pisa, Via Ghini 13, 56126 Pisa, Italy; 7) Department of Science of Nature and Territory, University of Sassari, Località Piandanna, Sassari, Italy; 8) The Wellcome Trust Sanger Institute, CB10 1SA, Hinxton, UK; 9) University of Pennsylvania, Philadelphia, PA, 10) CINVESTAV, Mexico City, Mexico; 11) University of San Martin de Porres, Lima, Peru; 12) Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil; 13) Instituto de Biologia Evolutiva (CSIC-UPF), Universitat Pompeu Fabra, Barcelona, Spain; 14) Vavilov Institute for General Genetics, Moscow, Russia; 15) Research Centre for Medical Genetics, Moscow, Russia; 16) The Lebanese American University, Chouran, Beirut, Lebanon; 17) University of the Witwatersrand, Johannesburg, South Africa; 18) Chettinad Academy of Research and Education, Chennai, India; 19) University of Arizona, Tucson, AZ; 20) University of Otago, Dunedin, New Zealand; 21) National Geographic Society, Washington DC, USA.

The search for a method that utilizes biological information to predict human’s place of origin has occupied scientists for millennia. Modern biogeography methods are accurate to 700 km in Europe but are highly inaccurate elsewhere, particularly in Southeast Asia and Oceania. The accuracy of these methods is bound by the choice of genotyping arrays, the size and quality of the reference dataset, and principal component algorithms. To overcome the first two obstacles, we designed GenoChip, a dedicated genotyping array for genetic anthropology with an unprecedented number of ~12,500 Y-chromosomal and ~3,300 mtDNA SNPs and over 130,000 autosomal and X-chromosomal SNPs carefully chosen to study ancestry without any known health, medical, or phenotypic relevance. We also 615 individuals from 54 worldwide populations collected as part of the Genographic Project and the 1000 Genomes Project. To overcome the last impediment, we developed an admixture-based Geographic Population Structure (GPS) method that infers the biogeography of worldwide individuals down to their village of origin. GPS’s accuracy was demonstrated on three data sets: worldwide populations, Southeast Asians and Oceanians, and Sardinians (Italy) using 40,000-130,000 GenoChip markers. GPS correctly placed 80% of worldwide individuals within their country of origin with an accuracy of 87%; for Asians and Oceanians. Applied to over 200 Sardinians villagers of both sexes, GPS placed a quarter of them within their villages and most of the remaining within 50 km of their villages, allowing us to identify the demographic processes that shaped the Sardinian society. These findings are significantly more accurate than PCA-based approaches. We further demonstrate two GPS applications in tracing the poorly understood biogeographical origin of the Druze and North American (CEU) populations. Our findings demonstrate the potential of the GenoChip array for genetic anthropology. Moreover, the accuracy and power of GPS underscores the promise of admixture-based methods to biogeography and has important ramifications for genetic ancestry testing, forensic and medical sciences, and genetic privacy.
1945F  
**AdmixKJump:** Identifying population structure in recently diverged groups.  
T. O'Connor1,2,3. 1 Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland; 2 Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland; 3 Department of Genome Sciences, University of Washington, Seattle, Washington.  
Correctly identifying population structure is important both to understand population history and to mitigate the potential for confounding in association analyses. Statistically, recent population divisions can be difficult to recognize, as there has not been substantial time for the groups to differentiate. Objective methods to identify recent population divisions are needed. ADMIXTURE has developed a cross-validation approach to select the correct number of K (i.e. clusters or putative populations), but how this statistic performs on recent population divisions with realistic simulations has yet to be evaluated. Also, alternative approaches from the informatics and statistical literature may be better suited to recent demographic events. I have implemented a new metric for admixture analysis, AdmixKJump, and compared it with the performance of the cross-validation statistic. I use a coalescent simulation framework based on parameters estimated from the Exome Sequencing Project to generate whole genome sequences with multiple populations. I vary the split time between populations to evaluate how accurately each method is in identifying the correct number of clusters over 50 replicates. With a sample size (n) of 50 for each of two populations, the admixture parameters themselves have almost no error at ~6KYA (i.e. ~1K years prior to the expansion). The cross-validation metric gains 100% power at about 14KYA, whereas AdmixKJump reaches 100% accuracy at 10KYA. I also find that the new measure has more power with smaller sample sizes (testing n=10, 20, .... 50), for instance n=30 is 100% at 12KYA for AdmixKJump. I then apply the new metric to the results of the Great Apes Project and 1000 Genomes Project. I find comparable results to the cross-validation approach, emphasizing the narrow improvement window of AdmixKJump for future improvements to fully capture the population structure in these data sets. In conclusion, I have developed a new objective approach to identifying population structure, which has more power than previous methods, especially with recent demographic events.

1946W  
**Distribution of genetic ancestry and candidate disease allele frequencies in Puerto Rico.** Y. Afanador1,2, I. Rivera1, W. Guibet1, LGDS. Consortium3, M. Yeager4, V. Washington5, J.C. Martinez-Cruzado1, T. Oleksyk1, 1 Biology, University of Puerto Rico, Mayaguez Campus, Mayaguez, PR; 2 Frederick National Laboratory for Cancer Research, Frederick, MD; 3 Biology, University of Puerto Rico, Rio Piedras Campus, San Juan, PR.  
In modern human populations, sensitivity to complex diseases is often determined by past demographic events, selection, and admixture. Puerto Ricans are an excellent model to understand this relationship as they acquired characteristics from three ancestral origins: African, European, and Native American. We are using a geographically distributed random sample of reference from each of the 78 municipalities of Puerto Rico collected by undergraduate students from a NSF-funded educational curriculum, the Local Genome Diversity Studies (LGDS). This study provides estimates for individual ancestry proportions across the island. Different admixture contributions carry different disease alleles that can account for prevalence of diseases that are on the top of the list of death causes in the Island: diabetes, kidney disease, heart disease, and cancer. Candidate genes for these diseases are being studied for their association with ancestry as well as the disease occurrence in the general population compared to the patients. For example, preliminary analysis of candidate gene polymorphisms for the end stage kidney disease (ESKD) patients from the island showed significant differences in frequencies of some disease-relevant variables, specifically in the NPHS2, ATP6, and ENPP1 genes when compared with the reference population, and the geographical distribution of the risk alleles shows a pattern consistent with the history of admixture across the island. These results demonstrate the importance of history and geography of admixture to the public health and personalized medicine decisions. Funding for this study came in part from NSF grant: DUE 1044714 TUES: Integration of Research and Undergraduate Education: Local Genome Diversity Studies in Puerto Rico.

1947T  
**Allelic Frequency Determination of Asthma-Related Single Nucleotide Polymorphisms and the Relation of Genetic Admixture in Asthma disease Prevalence among Puerto Ricans.** I. Rivera1, Y. Afanador1, C. Garcia1, W. Guibet1, E. Suarez2, J.C. Martinez-Cruzado1, T.K. Oleksyk1, Local Genome Diversity Studies. 1) Biology Dept, University of Puerto Rico, Mayaguez Campus, Mayaguez, PR; 2) Biology Dept, University of Puerto Rico, Ponce Campus, Ponce PR.  
Asthma is one of the most recognized complex human disorders characterized by an airway obstruction due to an exacerbated immune response. Currently, asthma is estimated to affect more than 235 million people worldwide with an annual cost of around $115 billion. In Puerto Rico asthma prevalence has been estimated to be 15% from 2000-2007, reaching an alarming 20% in 2003, making Puerto Ricans one of the populations with the highest asthma prevalence worldwide. Since asthma is more frequently seen in populations from African descent (around 12-14% Afro-Americans) and Puerto Ricans share on average 21% African ancestry, it has been suggested that an admixture factor might be involved. In order to evaluate this hypothesis, two sets of samples were collected: 1) a set of asthma-diagnosed patients; and 2) a systematically collected random sample set of individuals representing all 78 municipalities of the island: the Local Genome Diversity Studies cohort (LGDS). We assayed the allelic frequency of candidate asthma-related single nucleotide polymorphisms (SNPs) in patients and compared them to the general population reference samples from the LGDS cohort. In order to evaluate an additional admixture component that might contribute to asthma occurrence, we further analyzed the admixture proportions of African, European and Native American ancestry using a panel of ancestry informative markers (AIMs). Our results help validate admixture studies in genetically diverse Hispanic populations, as indicate the potential for novel ancestry-related candidate gene polymorphisms that correlate with asthma occurrence in Puerto Rican population. Funding for this study came in part from NSF grant: DUE 1044714 TUES: Integration of Research and Undergraduate Education: Local Genome Diversity Studies in Puerto Rico.

1948F  
**Evidence for interaction of population-specific EFHC1 alleles with genetic ancestry in juvenile myoclonic epilepsy.** R.L. Subaran1, J.M. Cerite1, W.C.L. Sterart1, D.A.G. Greenberg2. 1) Research Institute, Columbus, OH; 2) Nationwide Children's Hospital; 2) Columbia University, New York, NY.  
Misseense mutations in the EFHC1 gene have been reported to cause juvenile myoclonic epilepsy (JME) with high penetrance in patients reporting Hispanic ancestry. However, observations from other studies call into question the magnitude and scope of this effect. To help probe the relationship of EFHC1 to JME, we examine the frequency of these mutations in a group of Hispanics identified in New York City and in participants from non-Hispanic populations. We screened 117 healthy controls (60 Hispanic, 57 non-Hispanic of various ethnicities) for EFHC1 mutations purported to cause JME in Hispanics. To search for novel pathogenic mutations, we sequenced the exons of EFHC1 in a newly ascertained group of Hispanic JME patients. We compared our findings to the frequencies found in 1KGP. The EFHC1 coding mutations we found in our patients were also found in our controls except for a novel splice-donor mutation carried by a single JME patient. Of the five coding mutations previously reported to cause JME with high penetrance, we found three at appreciable frequencies in our non-Hispanic participants of the 1KGP. In general, EFHC1 coding mutations are not a major cause of JME in Hispanics. However, ascertaining specific Hispanic subpopulations on the basis of JME causes enrichment of EFHC1 alleles seen otherwise only in non-Hispanic populations, suggesting a hypersensitivity to disruption of EFHC1 in these Hispanic subpopulations. Importantly, our findings help shed light on the nature of this effect on a common disorder in a historically understudied population.
1949W

Evidence of archaic introgression into modern humans has accumulated in recent years. While most efforts to characterize the introgression process have relied on genome averages, only a small number of introgressive haplotypes have been shown to have an archaic origin after rejection of the alternative hypothesis of incomplete lineage sorting. Accurate identification of introgressive haplotypes is crucial both to characterize potentially functional consequences of archaic admixture and to quantify more precisely the genomic impact of archaic introgression. We perform two independent genomic scans for haplotypes of Denisova and of Neanderthal origin in a geographically diverse sample of complete genome sequences. These scans are based on the local sharing of polymorphisms and linkage disequilibrium, respectively. The analysis of concordance between the methods is then used to estimate the power and to compare demographic inference when performed using either all the data or just the genomic regions with no evidence of introgression. Moreover, we evaluate the extent to which Denisova haplotypes are observed in non-Melanoporean populations, and investigate whether the presence of such haplotypes is better explained by their persistence in the population since introgression or by more recent gene flow from Melanesians.

1950T
Admixture Estimation in a Founder Population. Y. Banda1, M. Kvale1, T. Hoffmann1, S. Hesselson1, H. Tang3, D. Ranalusung2, L. Walter2, C. Scafe2, P. Kwok1, N. Risca1, 1) Institute Human Genetics, University California San Francisco, San Francisco, CA, 2) Kaiser Permanente Northern California, Division of Research, Oakland, CA, 3) Department of Genetics, Stanford University, Stanford, CA.

Admixture between previously diverged populations yields patterns of genetic variation that can aid in understanding migrations and natural selection. An understanding of individual admixture (IA) is also important when conducting association studies in admixed populations. However, genetic drift, in combination with shallow allele frequency differences between ancestral populations, can make admixture estimation by the usual methods challenging. We have, therefore, developed a simple but robust method for ancestry estimation using a linear model to estimate allele frequencies in the admixed individual or sample as a function of ancestral allele frequencies. The model works well because it allows for random fluctuation in the observed allele frequencies from the expected frequencies based on the admixture estimation. We present results involving 3,366 Ashkenazi Jews (AJ) who are part of the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort and genotyped at 674,000 SNPs, and compare them to the results of identical analyses for 2,768 GERA African Americans (AA). For the analysis of the AJ, we included surrogate Middle Eastern, Italian, French, Russian, and Caucasus subgroups to represent the ancestral populations. For the African Americans, we used surrogate Africans and Northern Europeans as ancestors. For the AJ, we estimated mean ancestral proportions of 0.380, 0.305, 0.113, 0.041 and 0.148 for Middle Eastern, Italian, French, Russian, and Caucasus ancestry, respectively. For the African Americans, we obtained estimated means of 0.745 and 0.248 for African and European ancestry, respectively. We also noted considerably less variation in the individual admixture proportions for the AJ (s.d. = 0.02 to 0.05) compared to the AA (s.d. = 0.15), consistent with an older age of admixture for the former. From the linear model regression analysis on the entire population, we also obtain estimates of goodness of fit by r2. For the analysis of the AJ, the r2 was 0.977; for the analysis of the AA, the r2 was 0.994, suggesting that genetic drift has played a more prominent role in determining the AJ allele frequencies. This was confirmed by examination of the distribution of differences between observed and predicted allele frequencies. As compared to the African Americans, the AJ differences were significantly larger, and presented some outliers which may have been the target of selection (e.g. in the HLA region on chromosome 6p).

1951F

We apply a powerful haplotype-based model (described in Lawson et al. (2012)) to infer the population history of 410 individuals from 50 Native American groups, using data interrogated at >470,000 genome-wide autosomal Single-Nucleotide-Polymorphisms (SNPs). The model matches haplotype patterns among individuals’ chromosomes to infer which individuals share recent common ancestry at each location of the genome, an approach that has previously been demonstrated to increase power substantially over widely-used alternative approaches that consider SNPs independently. We apply this methodology to 1861 samples described in Reich et al. (2012), incorporating 263 additional samples from 32 relevant world-wide regions collated from other publicly available resources and currently unavailable data. We utilize these methodology and data in two ways. First, we infer intermixing (i.e. “admixture”) events among different Native American groups by identifying the groups that share the most haplotype segments. Using additional unpublished techniques, we determine the dates of these intermixing events, the proportions of DNA contributed, and the precise genetic make-up of the groups involved. These unique characteristics set this methodology apart from all presently available software, allowing us to place these mixing events into a clear historical context and thus identify the factors (e.g. the rise or fall of various Native American empires) that have contributed most to the genetic architecture of present-day Native American groups. Second, we match DNA patterns from each Native American group to a set of over 30 populations from Siberia and East Asia, describing each Native American group as a mixture of DNA from these regions. This enables us to shed light on the widely debated number of distinct migrations into the Americas during the initial colonization across the Bering Strait, comparing our results to previous inference from the literature. Our application demonstrates the power gained by using rich haplotype information relative to approaches that ignore this information.

1952W
Diversity of the Mexican Mestizo population using 18 X-STR. E. Ortiz1, G. Noris2, C. Santana3, M.A. Mera3, R. Gómez1, 1) Dep. de Toxicología, Cinvestav-IPN, Mexico City, Mexico; 2) Laboratorio BIMODI, Querétaro, Qro., Mexico; 3) Dep. de Biomedicina Molecular, Cinvestav-IPN, Mexico City, Mexico.

The use of the X chromosome STRs (X-STR) in population genetics, forensic studies and kinship testing have been amply used in recent years. The X-STRs shows intrinsic characteristics as lower mutation and recombination rates, and faster genetic drift due to smaller effective population size. In addition, population history and demographic factors have a deep impact on the population stratification. Previous studies have indeed highlighted the particularly complex genetic composition of the Mexican-mestizo population using Y chromosome STRs and mitochondrial DNA, however the population genetic studies using X-STR are still very limited. In order to assess the population stratification of the Mexican mestizo population we genotyped 200 unrelated women from central states of Mexico using 18 X-STR. Our results showed DXS10011 (k=32, PIC=0.94), DXS10134 (k=17, PIC=0.83), DXS10146 (k=22, PIC=0.84), and DXS10101 (k=21, PIC=0.90) were the most informative markers. Hardy-Weinberg departure was found in six out of 18 X-STR loci, with the most significant (P=0.00024) occurring at DXS10011. The use of the X chromosome STRs facilitates studies in general anthropological research.
Native American, European and African ancestry from genotype by sequencing in Argentinean populations.

Native American, European and African ancestry from genotype by sequencing in Argentinean populations. From our preliminary data, we have identified admixture on several thousand subjects from five different countries (Brazil, Argentina, Peru, Chile, and Colombia). We will focus on our sequencing data and compare it to previous studies.

In this work we aim to describe the demographic history and population structure of Argentina, by analyzing a collection of 2904 DNA samples from 15 dispersed regions in Argentina, from the Andean populations in the Northwest to the river lands of the Northeast. Argentina has history of broad cultural diversity, which included hunter-gatherers and agro-pastoralists to the expansion of the Inca Empire, the Spanish arrival, slave trade from Africa and the European migration between the mid XIXth and XXth centuries. Thus far, we have 390 genotypes defined from the Illumina Exome Array 250K and 89 genome-wide sequences from a genotype-by-sequencing strategy we developed, which covers 1.5% of the genome. We will continue to sequence more maps from our collection of 2904 participants from 15 Argentinean populations. From our preliminary data, we have identified proportions of Native American ancestry in the Argentinean cohort via local ancestry inference and found a correlation between latitude and proportion of Native American ancestry, where the highest proportion corresponds to the northernmost populations, (where it averages 72.0%) descending towards the center of Argentina (averaging 34.02%). We will focus on our sequencing data and compare it to full genomes of Native Americans, studying the genomic tracts of Native American ancestry for fine-scale examination of sub-continental structure.

Craniofacial morphometry has been one of the classical components of human physical anthropology. Recently there have been several studies on constructing 3-d models of the face and performing genetic associations. In this project, we bridge the two aspects of facial morphology and obtain interesting patterns with genetic ancestry.

The CANDELA (Consortium for the Analysis of the Diversity and Evolution of Latin America) project is exploring the effect of European-Native American admixture on several thousand subjects from five different countries (Brazil, Colombia, Mexico and Peru). From photographs, we construct a 3-d model of their faces with 36 landmarks. We also construct symmetrized distances, left-right deviations, and principal components of shape from these landmarks, each covering a different aspect of facial morphology. We observe these to be strongly correlated with ancestry. In general, European ancestry contributes to a larger facial shape, just as it contributes to taller height in our subjects. The noses and the lips show the largest amount of variation according to ancestry.

Following the style of classical anthropology, we independently phenotype our subjects on various traits using ordinal categories, such as the presence of epicanthic fold or mono-brow. We match these phenotypes with the quantitative distance measurements, and they show similar associations with European ancestry, with nose and lip traits again being the strongest. Finally, we genotype some of our Colombian samples on an Illumina 730K chip and perform a genome-wide association study on the traits. The GWAS implicates several genes such as NPAS2 and PDS3. As we genotype more samples from the other countries, we will be able to replicate the findings and probably find other associated genes as the sample size increases.
1956T
Population Structure and Genetic Diversity in a Population of 15,000 Patients from East Harlem, NY. G. Bellin1, D. Ruderrer1,2,3,4, E.A. Stahl1,2,5, J. Jeff2, Y. Lu2, R.J.F. Loos2,7, G. Oettgesen2, S. Purcell1,2,4, E. Bottinger6, E.E. Kenny1,2,4,6, 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Broad Institute, Cambridge, MA; 3) Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Center for Statistical Genetics, Icahn School of Medicine at Mt Sinai, New York, NY; 5) Institute for Personalized Medicine, Icahn School of Medicine at Mt Sinai, New York NY; 6) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mt Sinai, New York, NY.

New York City has historically been a significant point of entry for immigration into the United States and as a consequence is today populated by a highly stratified and diverse ethnic population. SeqScape, the DNA analysis software, reveals some of this diversity, but does not fully capture the variety of cultural groups, with complex and diverse demographic origins, foods and traditions, living in New York. Using genome-wide data, it is possible to detect such population structure which can both inform population history inference and result in better outcomes for medical genetics studies. We present a diversity of approaches for the analysis of fine-scale population structure in a population of 29,993 patients enrolled in the Icahn School of Medicine Biome Biobank Cohort (BioMe), of which ~13,500 have available Illumina Omni Express and Exome Chip data (~900K SNPs). BioMe comprises 34%, 47% and 19% participants with self-reported African-American (AA), Hispanic-Latino (HL) and European-American (EA), respectively, and is representative of the population of Northern Manhattan. We combined these data with both data generated from the 1000 Genomes Project and a unique database of genomic variation in over 4,000 individuals from diverse European, Middle Eastern, East Asian, African and Native American populations. Population genetic analysis using standard Principle Component Analysis (PCA), Principal Components (PC), and a Principal Component Analysis using Native American, European and African local ancestry haplotypes from AA and HL genomes, reveals diverse sub-continental structure in the BioMe cohort. In particular, we detect a large proportion of Ashkenazi Jewish and Eastern European ancestry in the BioMe EAs. We also perform an identity-by-descent (IBD) analysis and detect elevated cryptic relatedness in the AAs and HLs, which results in increased genetic tract sharing compared to EAs. For example, analysis of IBD haplotype sharing between any two less-than-fourth-degree relatives in our cohort indicates a larger percent of their genome is shared (~0.4% and 3.72%, for AA and HL, respectively), compared with the same analysis in EA (~0.12%).

1957F
Ultrafast and sample-aware local ancestry inference using population specific variants. R.P. Brown1, B. Pasaniuc1,2 1) Bioinformatics, UCLA, Los Angeles, CA; 2) Pathology and Laboratory Medicine, Geffen School of Medicine, Los Angeles, CA.

Inferring the ancestry at each genomic locus of recently admixed individuals (e.g. Latino Americans) plays a key role in medical and population genetic inferences from finding disease risk loci to inferring recombination rates. Current local ancestry methods are designed for genotyping arrays without utilizing linkage data and as data available today is very high, this approach is computationally expensive. In addition, existing methods do not make use of all the admixed samples when calling ancestry for a given individual and only rely on external reference panels used as proxies for the true ancestral populations. We present a fast and accurate method for inferring ancestry using population specific variants (PSVs) (i.e. variants identified by sequencing to be present in only one continental population) to infer local ancestry in sequence admixed genomes. We use the real 1000 Genomes data to find an abundance of such PSVs (e.g. an average of ~13.6 informative PSVs per Mb with ~28% coverage of the whole 109 in Africans) and model them within standard hidden Markov Models for local ancestry inference to achieve an ultra-fast and accurate approach. Our method includes an iterative framework that rebuilds its reference panels from the confidently called ancestry segments in the admixed individuals themselves to further boost accuracy while reducing bias introduced by the reference panels. Using simulations of Puerto Rican (Mexicans) from the 1000 Genomes, we show that our method achieves an accuracy of 0.93 (0.91) (quantified by the average r2 between inferred and simulated ancestry) as compared to 0.81 (0.90) achieved by existing methods. Most importantly, our approach is orders of magnitude faster than existing methods for full sequencing data (e.g. our method can infer local ancestry in 10,000 individuals in ~15 CPU days as compared to ~8 CPU years for existing methods). We also explore whether similar results can be achieved in real data and show that our approach yields comparable local ancestry to the provided calls in the real Puerto Rican and Mexican individuals from 1000 Genomes. We extend our approach to low coverage sequencing and show that accurate local ancestry can also be inferred at low coverage (e.g. r2 of 0.73 in our Puerto Ricans simulations at 4x coverage). Finally, we extend our method to GWAS-array genotyped individuals using a PSV tagging procedure that achieves accuracy similar to when full sequence data is available (e.g. 0.94 in African American simulations).

1958W
Y chromosomes in surname samples: insights into surname frequency and origin. F. Calafell, N. Solé-Morata, J. Bertranpetit, D. Comas. Instituto de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Barcelona, Spain.

In most societies, surnames are inherited through the paternal line, exactly as Y chromosomes are, with the exceptions of adoption, false paternity, and the inheritance of the maternal surname (as in the case of single mothers). The well-established phylogeography of the Y chromosome and the existence of fast-evolving STR markers implies that genotyping the Y chromosomes in samples of men bearing the same surname allows to i) count the number of founders of a surname and the frequency of their descendant in the current population and ii) pinpoint the remote origins of the founder of a surname. We have collected >2,500 samples from volunteers bearing one of 50 different Catalan surnames (http://cognoms.upf.edu/), in which we are genotyping 17 Y-chromosome STRs and a custom-designed set of 64 SNPs (typed in a single reaction with the OpenArray Real-Time PCR platform), with the following objectives: i) Discover and quantify the processes that drive surname frequency. Surname polymorphism (which can be measured from the number of founders detected from the Y chromosome diversity) can drive surname frequency, as was found in a sample of English surnames, but drift and natural selection (associated with high-status surnames) may also have a role. 2) Were the founders of surnames that are linguistically Arab or Hebrew North Africans or Jews themselves? 3) Were the founders of Germanic patronymic surnames of a different genetic origin form the rest of the population? In Catalonia, as in France, a frequent source of surnames are former first names of Germanic origin (Albert, Robert, Grav, ...). We will compare some of those to patronymic names in other populations (Danish, French, and a novel anonymous lineage). We will compare some of these to the objective and we have found that haplogroups that are more frequent in Germany or Gascony than in Catalonia were not overrepresented in the founders of the Alemay and Guasch surnames. Thus, these surnames may have originated as bynames not necessarily linked with the origins of the founders.

1959T
The Population Genetics of Sub-Saharan African Populations. M. Capredon1, J. Hussin1,2, J. Quinlan1,2, Y. Idaghdour1, T. de Maillard1, J.C. Grenier1, V. Bruat1, E. Gbeha1, J. Quinlan1,3, J.C. Capredon1, V. Bruat1, E. Gbeha1, J. Quinlan1,3 1) Pediatrics, CHU Sainte Justine, Fac Med, Univ Montreal, Montreal, Quebec, Canada; 2) Biochemistry, Faculty of medicine, University of Montreal, Montreal, Quebec, Canada; 3) School of public health, Faculty of medicine, University of Montreal, Montreal, Quebec, Canada.

Characterizing genetic diversity in African populations is critical for population genetics and personalized medical-genetic studies. We describe 14 populations, including four novel African datasets, two Western Africa (Cotonou and Zinwie) populations that are genetically very close, and two from Eastern Africa (Batwa pygmies and Kiga from Uganda). We compared these populations to ten published datasets from Africa Hapmap phase 3 (YRI, LWK, MKK) and HGDP (Bantu from Kenya, Bantu from South Africa, Biaka, Mandenka, Mbuti, San and Yoruba) across Africa. We also used a European dataset (Hapmap phase 3 CEU) as controls. Population structure analysis (PCA and admixture), revealed that the Kiga population is genetically very close to the LWK and to the Bantus from HGDP, while populations from Cotonou and Zinwie are clustering closely with the Yoruban and the Mandenka. Although the Zinwie population is geographically very close to Cotonou, clear differentiation was highlighted. Finally, the Batwa pygmies are substantially differentiated genetically from other populations. We observed that variability exists in recombination rates among African populations; sometimes this variability was associated with a geographiccline in Africa, but also was observed among more closely related populations (Yorubans from Nigeria versus Yorubans from Benin).
A scalable and effective local ancestry deconvolution algorithm for Latinos, G. Genovese1,2, A.L. Williams2, S.A. McCarroll2. 1) Stanford Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 2) Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

Genomes of admixed individuals are a mosaic of genetic segments inherited from different ancestral populations. Local ancestry deconvolution is the process of statistically inferring the ancestral source population of all sites in the genome of admixed individuals. Most local ancestry deconvolution algorithms are designed to work with SNP array genotype data and require reference panel data that are good representatives of each ancestral population. Obtaining reference panels may be especially challenging for Native American groups, and small or inaccurate panels have the potential to bias current methods. We developed a new algorithm (LATOOLS) for Latino local ancestry deconvolution which utilizes whole genome sequence data and does not require reference panels. The algorithm works in two steps. The first step identifies what we termed signature alleles: alleles that are present to an appreciable frequency in at least one ancestral population, but not in all ancestral populations for which the local ancestry deconvolution is computed. The second step uses these signature alleles to infer local ancestry in a given genome. Thus only sites that are fully informative for ancestry are incorporated into the model. While this model does not utilize all available information in its inference, our experimental results demonstrate that this approach is effective for Latinos. Furthermore, because of its relative simplicity, the algorithm is fast and scalable: its run time is linear in signature alleles count and cohort size. We show that in Latino samples from the 1000 Genomes project we are able to achieve, once several potential sources of bias are removed, and using step three of the model, an accuracy in the obtained local ancestry deconvolutions for the problem of admixture mapping the missing pieces of the human genome reference.

A scalable and effective local ancestry deconvolution algorithm for Latinos, A. Goldberg1, P. Verdu2, N.A. Rosenberg1. 1) Biology, Stanford University, Stanford, CA; 2) CNRS-MNHN, UMR7206 Eco-Anthro-phylogeny and Ethno-Biology, Paris France.

Complex social practices such as residence and descent rules, polygyny, and dominance relationships can produce sex-biased demographic histories in admixed human populations. Genetic signatures of sex-biased admixture have been empirically detected in samples from a variety of human populations throughout the world, usually without mechanistically modeling the full complexity of the sex-specific history. Expanding on the model of Verdu & Rosenberg (2011), we developed a model that mechanistically considers sex-specific admixture histories for autosomal DNA. Here, we extend our model to the sex chromosomes, allowing multiple source populations to contribute to the admixed population, potentially with varying contributions from male and female lineages across generations. Under the model, the X-chromosome is particularly informative about sex-biased admixture, as, unlike the autosomes, it can be used to identify which sex has a greater contribution to the admixed population. With no sex bias, the ratio of the expected X-chromosomal ancestry fraction to the expected autosomal ancestry fraction is one, but this ratio deviates from one with increasingly sex-biased admixture. The variance of the ancestry fraction is larger for a random X-chromosome sampled from a male from the admixed population rather than from a random female X-chromosome, in line with the decreased X-chromosomal population size for males compared to females. For admixture processes that are constant in time, we perform approximate Bayesian computation to infer sex-specific contributions from each source population. Considering autosomal DNA and the X-chromosome together, our approach can contribute to methods for inference of complex sex-biased admixture histories.

A two-sex model for the admixture history of a hybrid population: the X-chromosome, A. Goldberg1, P. Verdu2, N.A. Rosenberg1. 1) Biology, Stanford University, Stanford, CA; 2) CNRS-MNHN, UMR7206 Eco-Anthrophology and Ethno-Biology, Paris France.

Using this technique, we provide compelling evidence that the genetic differences among the Ari Blacksmiths and Cultivators -- including the Cultivators -- displaying varying proportions of African genetic ancestry. We tested whether sex phenotypes show an association with objectively measured individual genetic ancestry. Methods: The sleep of 101 Whites and 71 AA adults (mean age 58 ± 6 years; 57% female) was examined as part of the University of Pittsburgh SleepSCORE project. Polysomnography was used to assess average sleep duration, efficiency, Apnea-hypopnea index (AHI) and slow wave sleep (delta band of spectral EEG analyses). The Pittsburgh Sleep Quality Index (PSQI) was used to measure sleep quality. Biometric, psychosocial factors, health behaviors, and environmental factors were measured in all subjects. 1600 ancestry informative genetic markers were genotyped in all AAs which were used to infer overall individual African ancestry (IAA) by a maximum likelihood method. Multivariate tests were used to compare sleep phenotypes between Whites and AAs. Next, in AAs only, hierarchical linear regression was used to test whether IAA predicts sleep duration, efficiency, quality, architecture and AHI. Age, gender and BMI, education, income, use of sleep medication and antidepressants were used as covariates. All covariates were entered in step one, AIA (when not tested as a covariate) was entered in step two, and IAA was entered in step three of the model. Results: Sleep duration, efficiency and architecture differed significantly between Races. In AAs, IAA ranged from 10%-88% with a mean of 67%. Higher African genetic ancestry was associated with lower slow wave sleep in AAs (β(SE) = -4.6 (1.5); P = 0.002). This association persisted after all covariate adjustment. After adjusting for all relevant covariates individual African genetic ancestry explains 11% of the variation in slow wave sleep in AAs. Sleep duration, efficiency, quality or AHI was not associated with African genetic ancestry. Conclusion: Racial differences in slow wave sleep (but not other sleep phenotypes) appear to be explained partly by genetic variation related to continental ancestry in AAs.


Sleep characteristics vary by race. Compared to Whites, African Americans (AA) take longer to fall asleep, have lower sleep efficiency, lesser sleep duration and spend less time in the restorative slow wave sleep stage. The causes underlying these differences are not fully known but substantial heritability of sleep suggests some genetic underpinnings. We hypothesized that an overall genetic effect on sleep might be observed in AAs, and potentially varying proportions of African genetic ancestry. We tested whether sleep phenotypes show an association with objectively measured individual genetic ancestry. Methods: The sleep of 101 Whites and 71 AA adults (mean age 58 ± 6 years; 57% female) was examined as part of the University of Pittsburgh SleepSCORE project. Polysomnography was used to assess average sleep duration, efficiency, Apnea-hypopnea index (AHI) and slow wave sleep (delta band of spectral EEG analyses). The Pittsburgh Sleep Quality Index (PSQI) was used to measure sleep quality. Biometric, psychosocial factors, health behaviors, and environmental factors were measured in all subjects. 1600 ancestry informative genetic markers were genotyped in all AAs which were used to infer overall individual African ancestry (IAA) by a maximum likelihood method. Multivariate tests were used to compare sleep phenotypes between Whites and AAs. Next, in AAs only, hierarchical linear regression was used to test whether IAA predicts sleep duration, efficiency, quality, architecture and AHI. Age, gender and BMI, education, income, use of sleep medication and antidepressants were used as covariates. All covariates were entered in step one, AIA (when not tested as a covariate) was entered in step two, and IAA was entered in step three of the model. Results: Sleep duration, efficiency and architecture differed significantly between Races. In AAs, IAA ranged from 10%-88% with a mean of 67%. Higher African genetic ancestry was associated with lower slow wave sleep in AAs (β(SE) = -4.6 (1.5); P = 0.002). This association persisted after all covariate adjustment. After adjusting for all relevant covariates individual African genetic ancestry explains 11% of the variation in slow wave sleep in AAs. Sleep duration, efficiency, quality or AHI was not associated with African genetic ancestry. Conclusion: Racial differences in slow wave sleep (but not other sleep phenotypes) appear to be explained partly by genetic variation related to continental ancestry in AAs.
Patterns of genetic variation in populations of African ancestry observed in whole genome sequencing of 691 individuals from CAAPA. R. Mathias1, L. Huang1, T.D. O’Connor3, C. Vergara1, M. Taub1, A. Deshpande4, C.R. Gignoux6, N. Rafaels1, S. Shringarpure6, R. Torres1, J. Galanter2, R. Hernandez2, E.E. Kenny7, D. Locke7, W. Gruss2, K. Gietzen2, I. Ruczinski1, K.C. Barnes5, CAAPA Consortium. 1) Johns Hopkins University, Baltimore, MD; 2) University of Washington, Seattle, WA; 3) University of Maryland, Baltimore, MD; 4) Knome, Inc., Cambridge, MA; 5) University of California at San Francisco, San Francisco, CA; 6) Stanford University School of Medicine, Palo Alto, CA; 7) Icahn School of Medicine at Mount Sinai, New York, NY; 8) Illumina, Inc., San Diego, CA.

The Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) includes high coverage whole genome sequence (WGS) data (~30x depth) on 1,005 subjects of African ancestry and extends the patterns of variation catalogued in the Thousand Genomes Project (TGP) and Exome Sequencing Project (ESP) to a spectrum of populations representing a wide range of African ancestry. An interim data freeze (N=691) of CAAPA includes: 329 African Americans from 8 sites in the United States (Chicago, Atlanta, Baltimore/Washington, Nashville, New York City, Detroit, San Francisco, and Winston-Salem); 125 African Caribbean from three sites (Barbados, Jamaica, and Honduras); 212 African ancestry samples with a notable Latino component sampled from 5 Central and South American sites (Cartagena, Conde and Salvador in Brazil, Dominican Republic, and Puerto Rico); and 25 samples from Nigeria. The average Yoruba (YRI) component in CAAPA is 58%, ranging from <40% in the Puerto Rican samples to >80% in the Jamaican samples. Principal components and FST analysis reveals minimal differences between the 8 African American sites, of the 47.4 million (M) variants observed in CAAPA, a little more than half (~24M) are unique to CAAPA and not observed in TGP sequence data. 68.7% (~33M) have an MAF<1%, 13.8% (~6.6M) have an MAF 1-5% and only 17.5% (~8.4M) are common (MAF>5%). A dramatic 37% (~18M) sites represent private variation. Per genome, we observe ~4.1M variant sites and observe strong correlation between variation per genome with level of African ancestry (~r=0.97 with YRI content). Exome variation in CAAPA is similar to observations in the ESP. Relying on transcription factor binding motifs, we identified 5,812 enhancer predictions in the ESP. Enrichment analysis of predicted enhancer regions in CAAPA demonstrates that 33% represents novel African ancestry regulatory variation. 23K predicted enhancer site variants, ~3K sites in predicted promoter flanking regions, ~407K sites in predicted transcribed regions, ~13K sites in predicted promoter regions include TSS, and ~12K sites in predicted weak enhancer or open chromatin cis regulatory elements per genome. WGS data from CAAPA will provide an expansive understanding of genetic variation in populations of African ancestry and will be available as a public resource to the scientific community to improve our understanding of the role of genetic variation in complex disease.
1969F
Fossil free sequencing of archaic hominin metagenomes. B. Vernot, JM. Akey. Department of Genome Sciences, University of Washington, Seattle, WA.

To date, the genetic analysis of extinct archaic hominins has required the isolation and sequencing of ancient DNA obtained from fossilized remains, which is technologically challenging and limited by the number of available specimens. Here, we describe an alternative, fossil free paradigm, to sequence archaic genomes. The rationale of this approach is that ~3% of non-African genomes are estimated to have been inherited from Neanderthal ancestors. However, the precise introgressed sequences will vary among individuals. Thus, if enough individuals are analyzed, the various haplotypes of archaic sequence can be identified and stitched together to generate what we refer to as an ‘archaic metagenome’, as it consists of lineages from multiple archaic ancestors. To enable fossil free sequencing of archaic metagenomes, we developed a novel and computationally efficient statistic to identify introgressed DNA sequences, which is agnostic to the availability of an archaic reference sequence and can therefore facilitate the discovery of previously unknown archaic hominins, if such groups exchanged genes with modern humans. We rigorously evaluated the power and false discovery rate of our method through extensive coalescent simulations under a wide variety of demographic models and admixture scenarios. We applied our method to 379 European whole genome sequences, and identified a total of 5.2 Gb of putatively introgressed sequence, covering 1.1 Gb of the genome. These sequences are significantly enriched for matches to the high-coverage Neanderthal genome, supporting the hypothesis that they were inherited from admixture events with Neanderthal ancestors. Moreover, they are significantly depleted among coding regions, suggesting a fitness cost to hybridization. The recovered sequences also enable a variety of population genetics inferences to be made, such as effective population sizes, time of introgression, and number of archaic ancestors. In summary, we find that approximately 1/3 of the Neanderthal genome survives in modern humans and we anticipate that fossil free sequencing of archaic genomes will be a significant advance for the burgeoning field of paleogenomics, allowing genetic analyses that have heretofore not been possible.

1970W
No indication of Khazar genetic ancestry among Ashkenazi Jews. M. Metspalu1,2,4, D.M. Behar2,4, Y. Baran4, S. Rosset5, N. Kopelman5, B. Yuryusbayev1,2, A. Gladstein7, M.F. Hammer5, S. Tzur8, E. Halperin8,9, K. Skorecki10,11, R. Villems1,2, N.A. Rosenberg12; 1) Evolutionary Biology, Estonian Biocentre & Tartu Univ, Tartu, Estonia; 2) Molecular Medicine Laboratory, Rambam Health Care Campus, Hafia 31096, Israel; 3) The Blavatnik School of Computer Science, Tel Aviv University, Tel-Aviv 69978, Israel; 4) Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; 5) Porter School of Environmental Studies, Department of Zoology, Tel Aviv University, Tel Aviv 69978, Israel; 6) Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences, Ufa 450054, Russia; 7) ARL Division of Biotechnology, University of Arizona, Tucson, Arizona 85721, USA; 8) Department of Molecular Microbiology and Biotechnology, George Wise Faculty of Life Science, Tel-Aviv University, Tel-Aviv 69978, Israel; 9) International Computer Science Institute, Berkeley, California 94704, USA; 10) Ruth and Bruce Rappaport Faculty of Medicine and Research Institute, Technion-Israel Institute of Technology, Hafia 31096, Israel; 11) Estonian Academy of Sciences, Tallinn 10130, Estonia; 12) Department of Biology, Stanford University, Stanford, California 94305, USA; 13) Department of Integrative Biology, University of California Berkeley, 94720, USA; 14) these authors contributed equally.

The origin and history of the Ashkenazi Jewish population have long been of great interest. Most studies have concluded that the population derives its genetic ancestry from both Europe and the Middle East, and that it retains high genetic similarity to other Jewish groups such as the Sephardi Jews in Europe and Jewish communities in Northern Africa. It has recently been claimed, however, that a large part of the ancestry of the Ashkenazi population originates with the Khazars, a conglomerate of multi-ethnic, mostly Turkic-speaking tribes who consolidated into a powerful state just north of the Caucasus mountains between ca. 1,400 to 1,000 years ago. It has been difficult to explicitly test for Khazar contributions into Ashkenazi Jews, because it is not clear which extant populations can be used to represent modern descendants of the Khazars, and because the proximity of the southern Caucasus region to the Middle East makes it difficult to attribute any introgressed signal of Caucasus ancestry to Khazars. In particular, because it is not clear which extant populations can be used to represent modern descendants of the Khazars, and because the proximity of the southern Caucasus region to the Middle East makes it difficult to attribute any introgressed signal of Caucasus ancestry to Khazars.

The origin and history of the Ashkenazi Jewish population have long been of great interest. Most studies have concluded that the population derives its genetic ancestry from both Europe and the Middle East, and that it retains high genetic similarity to other Jewish groups such as the Sephardi Jews in Europe and Jewish communities in Northern Africa. It has recently been claimed, however, that a large part of the ancestry of the Ashkenazi population originates with the Khazars, a conglomerate of multi-ethnic, mostly Turkic-speaking tribes who consolidated into a powerful state just north of the Caucasus mountains between ca. 1,400 to 1,000 years ago. It has been difficult to explicitly test for Khazar contributions into Ashkenazi Jews, because it is not clear which extant populations can be used to represent modern descendants of the Khazars, and because the proximity of the southern Caucasus region to the Middle East makes it difficult to attribute any introgressed signal of Caucasus ancestry to Khazars. In particular, because it is not clear which extant populations can be used to represent modern descendants of the Khazars, and because the proximity of the southern Caucasus region to the Middle East makes it difficult to attribute any introgressed signal of Caucasus ancestry to Khazars.
As part of the largest Latin American initiative in population genomics and genetic epidemiology, we studied three Brazilian longitudinal populational cohorts: Salvador (n=1309), Bambuí (n=1442) and Pelotas (n=3736) from Northeast, Southeast and Southern Brazil respectively. We genotyped the Omni2.5 array for the 6487 individuals, the Omni5.0 array for 261 individuals and sequenced 30 complete genomes (average coverage: 42X). EPIGEN individuals show a very large extent of individual variability in ancestry proportions. While Native American ancestry was low (5-7% at population level, with no individual with > 30% of this ancestry), the three populations showed individuals with all possible combinations of African and European ancestry. At population level, African ancestry ranged from 14-19% in Pelotas and Bambuí to 51% in Bahia. Our unprecedented high resolution analysis of population structure of Brazilians in the context of worldwide variation shows that African and European ancestral subcomponents of Brazilians differ from African-Americans, reflecting the prevalent contribution from Mediterranean countries and African immigration from different geographic areas such as Mozambique and Angola. Our large and varied data set also allows us to identify genome-wide signal genes from the distribution of local chromosome ancestry. We are currently inferring the dynamics of the demographic admixture process in different parts of Brazil, as well as the time and mode of arrival to Brazil of clinically relevant mutations. First, we are identifying regions with strong signals of admixture, European, African or Native American ancestry, to identify candidate regions to be affected by Post-Columbian natural selection. Moreover, our high-quality sequencing data allowed us to identify between 3.6 M and 4.4 M of autosomal SNPs per each whole-genome sequenced individual, and find high levels of diversity and admixture of the Brazilian population allowed us to identify around 2.3 M of new autosomal SNPs, most of them rare. We are separating the African, European and Native American constituents of the 30 complete genome sequences and the 261 individuals sequenced, in order to analyze the distributions of different class of variants in function of their ancestry. The EPIGEN Initiative is also performing the 30 complete genomes to analyze the distributions of different class of variants in function of their ancestry. The EPIGEN Initiative is also performing several GWAS and admixture mapping studies on different complex traits, including longitudinal data. Funding: Brazilian Ministry of Health/FINEP.

1972F
The Genetic Structure and Admixture Analysis of Brazilian Populations: The Brazilian EPIGEN Initiative. H.C. Santos, A. Horimoto, A.C. Pereira, E. Tarazona-Santos, M.L. Barreto, B.L. Horta, M.F. Lima-Costa, M.T. Machado, T.M. Santos, J.M. Sanches, N. Esteban, W.C.S. Magalhães, M.R. Rodrigues, F.S.G. Kehdy, The Brazilian EPIGEN Consortium. 1) Laboratory of Genetics and Molecular Cardiology, Heart Institute, Medical School of University of São Paulo; 2) General Biology Department, Federal University of Minas Gerais, Brazil; 3) Instituto de Saúde Coletiva, Federal University of Bahia, Brazil; 4) Universidade Federal de Pelotas, Brazil; 5) Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil. Brazilians form one of the most heterogeneous populations in the world, the result of interethnic crossings between peoples from three continents: the European colonizers, African slaves and the autochthonous Native Americans. To analyze patterns of human genetic variation in Brazilian populations, we used data from 370,539 shared SNPs from the EPIGEN-Brazil Project genotyped by Omni2.5 Illumina array (6,487 individuals from three Brazilian cohorts: Pelotas, Southern Brazil, 3,736 individuals, Minas Gerais, South East Brazil, 1,442 individuals and Bahia, north east Brazil, 1,309 individuals). 6,889 individuals from 66 populations from HapMap Project (CEU, YRI, ASW, TSI, MEX and LWK) and 43 individuals from 9 American populations from HGDP (Karitiana, Surui, Pima and Maya). Principal components and individual admixture analysis (by ADMIXTURE software) were performed to these populations to identify structure, ancestry percentages and a minimum set of SNPs needed to capture the admixture components of Brazilian populations. As expected, Brazilian samples fell between Africans, Europeans and Mexicans. Furthermore, the mean percentage of individual European, African and Native American ancestry, respectively were 0.773, 0.185 and 0.042 for Pelotas cohort, 0.798, 0.142 and 0.059 for Bahia cohort and 0.435, 0.505 and 0.060 for Bahia cohort. From 370,539 analyzed SNPs, we found a minimum number of 256 SNPs able to capture 91%, 93% and 78% of respectively African, European and Native American ancestry components in the three Brazilian populations. The high accuracy of these sets of SNPs suggests this population as a good tool for admixture mapping studies. More studies about intracontinental components that contributed to the Brazilian population formation are needed and will be performed. Funding: Brazilian Ministry of Health/FINEP.

1973W
Population stratification detection and correction in rare variant collapsing methods using principal component analysis. J.R. Wallace, C.B. Moore, M.D. Ritchie. 1) Center for Systems Genomics, The Pennsylvania State University, University Park, PA; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN. Principal Component Analysis (PCA) has often been used in genome-wide association studies (GWAS) to detect population stratification and prevent increased type I errors. With the proliferation of Next Generation Sequencing (NGS) technology, rare variant collapsing methods have gained popularity. However, applying PCA to rare variant collapsing methods has not been well studied, and the effectiveness of adjusting for population stratification using principal components (PCs) on rare variants is largely unknown. To explore population stratification correction in rare variant data, we collapsed the low frequency (< 5% MAF) variants in the 1000 Genomes Project Phase 1 data based on Entrez gene boundaries using BioBlitz, a bioinformatics tool for automatically binning rare variants into biologically relevant bins (genes, pathways, regulatory regions, etc.) based on prior biological knowledge gleaned from various public data sources such as Entrez, Gene Ontology (GO), and Protein Families database (PFAM). These bins can then be evaluated with any number of popular rare variant collapsing statistical methods. We analyzed the data for each pairwise combination of the 14 populations available in the 1000 Genomes data and found dramatic stratification and natural clustering of populations into continental groups. We then examined multiple approaches to construct PCs using different subsets of the genetic data. To compare approaches, we defined a normalized distance metric between sets of PCs as well as notions of correctness for a given stratification and predictive power without given stratification. We illustrate these concepts with respect to the 1000 Genomes data and show how similar concepts can be applied to a natural dataset in the course of analysis. We found that for populations close in ancestral history, rare variants should be excluded from analysis regions with high frequencies of PCs. Although this method can disguise other hidden stratification, identifying and properly correcting for stratification remains an important issue; using the techniques described, we demonstrate solutions that identify and correct the ancestry stratification as well as stratification along sequencing technology.

1974T
Surveying European and West African Population Structure Using >2,300 Samples with Spatial Information. Y. Wang, K. Noto, J.B. Byrnes, R.E. Curtis, N.M. Myres, M.J. Barber, J.M. Granka, C.A. Ball, K.G. Chahine. AncestryDNA, San Francisco, CA. Population structure arises as a consequence of the interplay between geography, genetic drift, and gene flow. Knowledge of human population structure is fundamental to understanding how demographic history has shaped human genetic diversity. In this work, we investigated the genetic structure of European and West African populations using high-density SNP data. First, we analyzed ~2000 European individuals genotyped at >700,000 SNPs using principal component analysis (PCA), spatial ancestry analysis (SPA), and admixture-model-based method (ADMIXTURE). Despite an overall high level of genetic similarity, we observed significant patterns of genetic variation among European populations, ranging from the regional level (e.g., Northern European vs. Southern European and Eastern European vs. Western European) to the local level (e.g., Iberian vs. Italian and English vs. Irish). We then conducted similar analyses on ~330 individuals sampled from nine West African countries (Senegal, Mali, Ivory Coast, Ghana, Benin, Togo, Nigeria, Cameroon and Congo). Our results revealed a strikingly significant structure at the country level, despite geographical proximity. Comparing the three approaches, we found that SPA has better power in predicting the population of single origin individuals, given an accurate reference panel of reasonable size. We further applied SPA to identify SNPs showing large gradients in allele frequency, which can be used to tag candidate regions under natural selection. In addition, we set two sets of ancestry informative markers (AIMs) that carry substructure information for European and West African populations, respectively. Lastly, we introduced a new tool for visualizing the pattern of genetic variation using a hierarchical clustering (HC) based large cohort (>5,000) of African American individuals and the European and West African reference panels. Our results shed new light on the population migration history of African Americans.
Using evolutionary profiles to better-inform model organism selection for human disease research. A.D. Baxevanis, E.K. Maxwell, C.E. Schmitzler, A.D. Nguyen, R.T. Moreland. Genome Technology Division, Division of Intramural Research, National Human Genome Research Institute, NIH, Bethesda, MD.

While the standardization of methods for studying human diseases in traditional animal models has yielded many clinically actionable results, it has effectively narrowed the breadth of species in which we choose to look for insights. The recent expansion of whole-genome sequence data available from a diverse array of animal lineages provides an opportunity to investigate the feasibility of using non-traditional model organisms to advance our understanding of human diseases. Cases in which traditional animal models have led to conclusions that are not applicable to humans are becoming more commonplace, and the concern that this may be a growing problem calls for a re-evaluation of how appropriate models are selected for different disease classes. To that end, we have used a comparative genomics approach that encompasses a wide range of animals across the metazoan tree to determine which organisms could serve as viable models for studying various classes of human diseases. We show that some emerging non-bilaterian model organisms have surprisingly high proportions of human disease gene homologs despite their great evolutionary distance from humans; these organisms may confer advantages as animal models in terms of their ease of use, short generation times and cost-effectiveness. Conversely, while it has been previously shown that the genes implicated in the causation of most human diseases are of ancient origin, our results indicate that some disease classes involve a significantly large proportion of genes that appear to have emerged relatively recently within the Metazoa. This disease class discovery, having a more recent evolutionary history, may have different implications for the causation and treatment of these diseases including those affecting blood coagulation and lipid homeostasis. For example, several of the promoter regions in the blood coagulation cascade convey highly conserved CRMs that are recurrently mutated in bleeding disorders. In primates, the CRMs implicated in disease point to pathways involving drug detoxification. Deeply conserved lineage-specific CRMs were also found in close proximity to liver and blood lipid-disease loci identified by genome-wide association studies (GWAS). Together, our data demonstrate that the rapid evolution of human combinatorial transcription factor binding can be used to identify and prioritize the functional study of pathologic regulatory mutations.

A Model-Based Analysis of GC-Biased Gene Conversion in the Human and Chimpanzee Genomes. J.A. Capra1, M.J. Hubisz2, D. Kostka3, K.S. Pollard4, A. Slepé5, 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 3) Developmental Biology and Computational & Systems Biology, University of Pittsburgh, Pittsburgh, PA; 4) Institute for Human Genetics and Division of Biostatistics, University of California, San Francisco, CA.

GC-biased gene conversion (gBGC) is a recombination-associated process that favors the fixation of G/C alleles over A/T alleles. In mammals, gBGC is hypothesized to contribute to variation in GC content, rapidly evolving sequences, and the fixation of deleterious mutations, but its prevalence and general functional consequences remain poorly understood. gBGC is difficult to incorporate into models of molecular evolution and so far has primarily been studied using summary statistics from genomic comparisons. Here, we introduce a new probabilistic model that captures the joint effects of natural selection and gBGC on nucleotide substitution patterns, while allowing for correlations along the genome in these effects. We implemented our model in a computer program, called phastBias, that can accurately detect gBGC tracts in large mammalian genomes. When applied to real primate genome sequences, phastBias predicts gBGC tracts that cover roughly 0.3% of the human and chimpanzee genomes and account for 1.2% of human-chimpanzee nucleotide differences. These tracts fall in clusters, particularly in subtelomeric regions; they are enriched for recombination hotspots and fast-evolving sequences; and they display an ongoing fixation preference for G and C alleles. They are also significantly enriched for disease-associated polymorphisms, suggesting that they contribute to the fixation of deleterious alleles. The gBGC tracts provide a unique window into historical recombination processes along the human and chimpanzee lineages. They supply additional evidence of long-term conservation of megabase-scale recombination rates accomplished by rapid turnover of hotspots. Together, these findings shed new light on the evolution, functional, and disease implications of gBGC. The phastBias program and our predicted tracts are freely available.

Conserved directional crossovers in human and chimpanzee genomes. S. Merkle, M. Gonzalez-Porta, M. Carluccio, K.N. Chessman, A. Faure1, A. Funnell2, A. Goncalves3, C. Kutter4, M. Lukk2, S. Menon3, W.M. McLaren1, K. Srebotnik1, S. Watt2, J.C. Marion1, D.T. Odom1, F. Flicek1, 1) Genetics and Genome Biology, SickKids Research Institute, Toronto, ON; 2) INSERM U915, TAGC, Aix-Marseille University, Marseille, France; 3) University of Cambridge, Cancer Research UK, Cambridge Institute, Robinson Way, Cambridge CB2 0RE, United Kingdom; 4) European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom; 6) Department of Molecular Genetics, University of Toronto, Canada 101 Collingwood East Tower, Toronto, ON, M5G 1L7, Canada; 7) School of Biotechnology and Biomolecular Sciences, University of New South Wales, NSW 2052, Australia.

In order to test whether evolutionary conservation of combinatorial protein-DNA interactions gives insight into human gene regulatory function and disease, we experimentally determined transcription factor (TF) binding locations of four master regulatory TFs (ONECUT1/HNF6, FOXA1/HNF3A, HNF4A, and CEBPA) in the livers of human, macaque, mouse, rat and dog. Approx., three fourths of the TF binding events for these four TFs resided in heterotypic clusters, which we designated as cis-regulatory modules (CRMs). Less than half of the identified human liver CRMs could be detected in orthologous regions from one or more of our study species. We found that CRMs conserved in human and at least one non-primate to be disproportionately enriched in regions found in recurrently mutated genes, including those affecting blood coagulation and lipid homeostasis. For example, several of the promoter regions in the blood coagulation cascade contain highly conserved CRMs that are recurrently mutated in bleeding disorders. In primates, CRMs implicated in disease point to drug detoxification. Deeply conserved lineage-specific CRMs were also found in close proximity to liver and lipid-disease loci identified by genome-wide association studies (GWAS). Together, our data demonstrate that the rapid evolution of human combinatorial transcription factor binding can be used to identify and prioritize the functional study of pathologic regulatory mutations.

Functional characterization of Toll-like receptor signaling pathways in primates. J.F. Brinkworth1,2, J.N. Kohn2,4, J. Boulais1,2, J.C. Grenier1, R.E. Lanford1, Z.P. Johnson3, L.B. Barreiro2,4, 1) Centre Hospitalier Universitaire Sainte-Justine Hospital Centre de Recherche; 2) Department of Pediatrics, University of Montreal, Quebec, Canada; 3) Yorkes National Primate Research Center, Atlanta, GA, United States; 4) Graduate Program in Neuroscience, Emory University, Atlanta, Georgia, United States; 5) Department of Biochemistry, University of Montreal, Quebec, Canada; 6) Department of Biochemistry, University of Montreal, Quebec, Canada. Supported by the Fonds de Recherche du Québec - Santé (FRQS) and the Canadian Institutes of Health Research (CIHR, Grant # TGF-96109) (JFB).
1979W
Accelerated evolution of primate-specific miRNAs in the human genome. M. Lopez-Valenzuela1, N. Petit-Marty1, A. Navarro1,2, Y. Espinosa-Parrilla1. 1 Institute de Biologia Evolutiva (IBE, Universitat Pompeu Fabra-CSIC), Barcelona, Catalonia, Spain; 2 Instituto Nacional de Bioinformática (INB), Barcelona, Catalonia, Spain; 3 Instituto Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

miRBase miRNAs are a class of small non-coding RNAs with a prominent role in the genome regulatory networks. miRNA genes are continuously being added to metazoan genomes allowing the creation of lineage-specific miRNAs. The differential regulation provided by these lineage-specific miRNAs may contribute to shape different phenotypes in species with genomes of similar protein-coding content. The rise of primate lineage is marked by an outstanding emergence of new miRNA families, but how natural selection has acted on them is still not known. We studied the action of natural selection on human miRNAs applying a likelihood ratio test based in comparing substitution rates in the sequences of the miRNAs themselves with those in ancestral repeat (AR) sequences that are used as the reference for neutrality. The test can detect positive selection acting upon a small number of sites and distinguish it from relaxation of purifying selection. Primate PhyloP scores for each nucleotide in the studied miRNAs were also calculated as a complementary and alternative method. Comparison of 1,523 known human precursor miRNA sequences (miRBase release 18) with the reference genomes of chimpanzee and rhesus macaque showed nucleotide differences for 555 human miRNAs. The great majority of these miRNAs did not reach significance in the likelihood ratio test indicating that most sites in human miRNAs are either conserved or evolving neutrally. Nevertheless, two miRNAs, hsa-mir-518a-1 and hsa-mir-5939, escaped markedly this tendency and presented the signature of accelerated evolution by positive selection. Interestingly, the primate-specific fraction of the analyzed miRNAs (79%) presented a higher proportion of significant nucleotides than the rest of miRNAs (p=0.00001) and showed lower average conservation index in the PhyloP analysis indicating that this group of primate-specific miRNAs could constitute a group of fast evolving miRNAs. We also found that, according to the TargetScan algorithm, these accelerated miRNAs tend to have less targets and that functions related to their target genes are similar to those of other families. The study of the miRNAs observed in the primate lineage could allow an understanding of the histone regulation of addiction behavior, which show high derived variant frequencies in African ancestry populations (Luhya, Masaai, African American, and Yoruba). Of particular interest are two polymorphisms (rs9467667 and rs9379817) which sit in the regulatory region of the HIST1H1C in a GABA specific addiction cluster, which show high derived variant frequencies in Chinese populations but which is nearly fixed for the ancestral variant in African ancestry populations. These variants may play a role in epigenetic histone regulation of alcohol sensitivity behavior.

1980T
Addiction Drugs Cluster Functionally in the Genome with Concomitant Variation in Human Populations. L. Jackson1, Y. Liu2, A. Tozeren1. 1 Biomedical Engineering, Science, and Health System, Drexel University, Philadelphia, PA, USA.

Understanding the common genetic mechanisms that underlie dopamine, opioid, and GABA addiction is of significant interest to human genetics. We seek to use biologically relevant addiction genes as the starting pool to identify whether genes involved in addiction cluster together into loci of functional addiction control. We ask whether genes involved in addiction physically cluster in the genome, if there is a functional pattern to this clustering and whether such clusters show polymorphic variation in HapMap populations. Genes with biological relevance related to addiction (N=587) were gathered using NCBI Gene and then physically mapped onto the human genome using computational tools. Clusters of genes were identified using specific threshold criteria. These clusters were then surveyed to determine their drug addiction class membership. Polymorphisms in each of these regions were identified for each of the 11 HapMap populations. To test for deviations from neutral demographic expansion at these sites by using autosomal comparison regions and mapped within-cluster variation. Our analyses found nine addiction gene clusters with three clusters containing GABA specific addiction genes and six clusters containing addiction genes involved with all surveyed classes. Furthermore, analysis of the polymorphisms underlying these clusters show patterns consistent with non-neutral GABA specific evolution in Chinese populations (sampled in either Beijing or Denver), and with non-neutral evolution at generalist addiction genes in African ancestry populations (Luhya, Massai, African American, and Yoruba). Of particular interest are two polymorphisms (rs9467667 and rs9379817) which sit in the regulatory region of the HIST1H1C in a GABA specific addiction cluster, which show high derived variant frequencies in Chinese populations but which is nearly fixed for the ancestral variant in African ancestry populations. These variants may play a role in epigenetic histone regulation of alcohol sensitivity behavior.

1981F
Critical Assessment of Coalescent DNA Simulators in Modeling Recombination Hotspots. T. Yang, H.W. Deng, T. Niu*. Dept of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA.

Coalescent-based simulation plays a pivotal role in understanding population evolutionary models and demographic histories, as well as in developing new analytical methods for genetic association studies. A plethora of coalescent simulators are developed, but it often remains challenging to select the most appropriate program. We extensively compared performances of five most-widely used ones - Hudson’s ms, msHOT, MaCS, Simcoal2, and fastsimcoal, to provide a practical guide considering three crucial factors: (i) speed, (ii) scalability and (iii) hotspots position accuracy. Hudson’s ms, the most popular simulator has very robust performance under standard coalescent but lacks the ability to simulate sequences with recombination hotspots. We highly recommend ms in absence of hotspots. msHOT has compensated the deficiency of ms incorporating crossover and gene conversion hotspots at arbitrary location and intensity. Simcoal2, based on discrete-generation coalescent, could simulate more complex demographic scenarios, but is comparatively slow. MaCS and fastsimcoal, both adopting a fast sequential Markov coalescent model to approximate standard coalescent, are much more computationally efficient whilst keeping salient features of msHOT and Simcoal2, respectively. Our extensive simulations demonstrated that MaCS and fastsimcoal have significant advantages over other programs for a spectrum of demographic scenarios. We also evaluated accuracy of hotspot positions of the simulated data by LDhat 2.2 rhomap package, sequenceLDhat and haploview. We found that fastsimcoal has the best performance. In conclusion, while ms remains an excellent choice for general scenarios, MaCS and fastsimcoal are much more scalable and suitable for the complex demographic histories and diverse DNA sequence structures. *Corresponding Author.

1982F
The genome-wide distribution of gene conversion, cross-overs, and de novo mutations events in Western Chimpanzees. O. Venn1, I. Turner1, T. Mathieson1,2, N. de Groot3, G. McVean1,2. 1 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) University of Oxford, Department of Statistics, Oxford, OX1 3TG, United Kingdom; 3) Biomedical Primate Research Centre, Department of Comparative Genetics and Refinement, Rijswijk, 2288, Netherlands.

A wide spectrum of processes including cross-over, gene conversion, mutation and DNA repair impact genomes from the level of single nucleotide changes to multiple megabases. Inter-species comparisons of their genome-wide distributions may potentially be informative about the relative impact of each process and their evolution. The analysis of genetic variation measured through high-throughput population sequencing is a potentially powerful approach to study such processes. However, a major challenge in interpreting this information is the detection of erroneous, mis-assembled, or incorrectly genotyped genetic variants.

To characterize the spectrum of genome changes occurring in Western chimpanzees (Pan troglodytes verus) we sequenced an extended three-generation pedigree (3 founders, 3 F1, 3 F2) to ~30x across individuals. This family structure maximizes transmission information for the detection of rare events: 7/8 of the founder genomes is transmitted at least once, events arising in the F1 parent are naturally validated in the F2s, detected events can be assigned parental origin, and 99.2% of segregating sites can be phased unambiguously. At fine-scales, we describe the properties of de novo mutation events with respect to parental age and the impact of local population size. We estimate a sex-averaged de novo mutation rate of 1.38 × 10⁻⁷ bp per meiosis. To characterize meiotic recombination, we infer ~400 recombination hotspots at arbitrary location and intensity. Simcoal2, based on discrete-generation coalescent, could simulate more complex demographic scenarios, but is comparatively slow. MaCS and fastsimcoal, both adopting a fast sequential Markov coalescent model to approximate standard coalescent, are much more computationally efficient whilst keeping salient features of msHOT and Simcoal2, respectively. Our extensive simulations demonstrated that MaCS and fastsimcoal have significant advantages over other programs for a spectrum of demographic scenarios. We also evaluated accuracy of hotspot positions of the simulated data by LDhat 2.2 rhomap package, sequenceLDhat and haploview. We found that fastsimcoal has the best performance. In conclusion, while ms remains an excellent choice for general scenarios, MaCS and fastsimcoal are much more scalable and suitable for the complex demographic histories and diverse DNA sequence structures. *Corresponding Author.
1983T
Moving backwards in time from present-day sequences: the paradox of sequential founder events. M. Jeanpierre. Department of genetics, Hôpital Cochin, Paris, France.

It is difficult to unravel the history of ancient populations, and events identified as single founder migrations may actually turn out to be a sequential series of founder events, with a small number of initial settlers gradually joined by an increasing number of related individuals of the same origin. The modeling of sequential founder events is an important issue, because each genetic variant is studied as a member of a sequence, and it is only possible to filter out these adjacent variations if we know the distribution of identical-by-descent segments. When modeling of the decay of haplotype sharing, simple geometric structures that can be described unambiguously in mathematical terms can provide an algebraic framework for analyses of the forces shaping the genealogy of a single allele. The challenge is to develop simplifications with a minimal loss of information, because the precision of the reconstruction depends on the dimensionality of the graph. The definition of blocks as physical entities, with clear borders, as for objects in the physical world, results in an apparent simplification, but is not really helpful because segments identical by descent are statistical entities with an ephemeral existence. Paradoxically, a hierarchical model, in which clusters are considered as border-less elements, gives a better representation of ancestral sequences, its precision steadily increasing with sample size. Large samples are not uncommon, with some containing more than 100,000 individuals, so this relationship between sample size and precision matters. The weakness of low-dimensional contingency tables may be one of several explanations for the poor reproducibility of association studies. This theoretical approach may be illustrated by the non random distribution of genetic variants across the X chromosome. The clustering of variations defining ephemeral units of selection makes sense from a genetic point of view, because selection operates on visible physical characters, not on numerous, infinitesimally small genetic elements.

1984F

The genetic diversity of a population reflects the rate of mutation (µ) creating new alleles, and recombination events (c) that redistribute these alleles among homologous chromosomes to form a variety of haplotypes. Importantly, genetic diversity also reflects the historical effective size (Ne) of a population, since in larger populations more diversity may arise every generation and surviving variants persist for longer, thus resulting in greater population diversity. Quantitatively, the population parameters θ=4Neµ and θc=4NeC define the mutational and recombinational diversity, respectively. Combining different θ estimates that reveal different aspects of population genealogies has led to statistical tests that are sensitive to selection events and/or demographic history. Here, by computer simulations, we explore different estimates of θ and compare their sensitivity to detect recombination and gene conversion under a standard neutral population model at constant population size, as well as following population bottleneck and demographic growth. We compare θ estimates obtained from counting past recombination events (R) with those evaluated from the extent of linkage disequilibrium (LD) along the genome sequence. While pLD is practically insensitive to gene conversions, thus monitoring only reciprocal crossover events, pR monitors both recombination and gene conversion. Interestingly, following a population bottleneck these p estimates recover at different rates, as observed with different θ such as Tajima’s and Watterson’s estimates. Both p estimates were also computed for the autosomes of all HapMap3 populations. An excellent correlation was observed when individual chromosomal estimates were compared, not only between our two methods, but also with LDHat as well as with recent pedigree estimates of the recombination rate.

We expect that combining different recombinational diversity estimates and, especially, in combination with different mutational diversity estimates, should open up new ways to infer demographic as well as population genetic history, including events of natural selection.

1985W
Analysis of linkage disequilibrium associated with Southeast Asian Ovalocytosis (SAO). M.K. Thompson, J.A. Wilder. Northern Arizona University, Flagstaff, AZ.

Southeast Asian Ovalocytosis (SAO) is a type of hereditary elliptocytosis that confers protection from malaria-causing parasites, including both Plasmodium falciparum and P. vivax. This trait is caused by a 27-base pair deletion, or a 9 amino acid deletion, in the Band 3-encoding SLCA41 gene. While SAO provides protection from severe manifestations of malarial parasitism when heterozygous, the trait is lethal when homozygous. This pattern suggests it is maintained as a balanced polymorphism. Previous studies have found substantial short-range decay in linkage disequilibrium (LD) associated with the SAO-causal mutation, suggesting a relatively ancient age of the trait. Here we extend this analysis by targeted resequencing of loci that span 500 kilobases around the causal mutation in 52 SAO-carriers from Indonesia and Thailand. These data will allow us estimate the decay of LD over a relatively large chromosomal region, spanning a highly heterogeneous recombination environment. These results will refine our understanding of the evolutionary history of Southeast Asian Ovalocytosis.

1986T
Meliotic gene conversion in humans: rate, sex ratio and GC bias. A.L. Williams1, G. Genovese1, T. Dyer2, N. Patterson1, J. Blangero2, D. Reich1,3.

the T2D-GENES Consortium. 1) Broad Institute, Cambridge, MA; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) Harvard Medical School, Boston, MA.

Gene conversions are short, 20-300 bp segments copied from one homologous chromosome to another during meiosis. While related to crossover, gene conversions are distinct in that they affect a much shorter region, are more frequent, and show bias towards G or C allele transmissions. To date, there has been no genome-wide study of de novo gene conversion and no direct inferences about their localization, sex differences, and the impact of GC bias on the genome. We examined SNP array data from 190 individuals in 16 three-generation pedigrees that are informative for 42 meioses (21 paternal, 21 maternal) and 4.15×10⁸ sites. Our study design uses three-generation pedigrees and requires that a gene conversion received by a second generation child is also transmitted to a third generation grandchild. This ensures that the identified gene conversions are either real or the result of at least two genotyping errors at the same site.

We identified 33 putative gene conversions and obtained a genome-wide de novo gene conversion rate estimate of 8.0x10⁻⁶ per bp per generation—a rate consistent with sperm typing and LD-based studies. We validated these gene conversions using whole genome sequence data for a subset of the genotyped samples, with genotype calls available for 19 of the 33 sites. Of these 19 sites, 18 had genotypes consistent with the SNP chip data. The single mismatching site is ambiguous as to the error source and appears to be a sequencing artifact. These results suggest a low false-positive rate and validate our study.

The gene conversions are significantly enriched in recombination hot spots, with 10/33 events in regions with recombination rate ≥10 cM/Mb (P=1.1x10⁻⁸). Male gene conversions predominantly localized to telomeres while females transmitted 1.54x more gene conversions than males. These results further validate our methodology and are consistent with the mechanism for forming gene conversions being similar to that of crossovers.

We observed extreme GC bias in allelic transmissions, with 23/31 sites heterozygous for AT and GC alleles transmitting a G or C allele (P=5.3x10⁻⁸). This 3-fold enrichment of GC transmissions suggests that gene conversion substantially alters the allelic make up of the genome. We are examining a 2-fold larger dataset to further validate this and other findings.
1987F
Linkage disequilibrium and haplotype blocks determined by the analysis of 250K SNPs in three quilombo remnants communities. E. Simões, C.T. Mendes-Junior, D.M. Salvanha, R.Z.N. Vênico, L.M. Garrido, H. Krieger, A.L. Simões, C.T. Mendes-Junior. 1) Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto-SP, Brazil; 2) Departamento de Matemática e Computação, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto-SP, Brazil; 3) Departamento de Parasitolgia, Instituto de Ciências Biomédicas, Universidade de São Paulo, 05058-000, São Paulo-SP, Brazil; 4) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto-SP, Brazil. The non-random association between alleles of different loci characterizes what is called linkage disequilibrium (LD) between them. The LD extent in human populations can be influenced by many factors, such as recombination rate, demographic features and evolutionary events. The aim of this study was to describe the LD patterns of four Brazilian populations and correlate these patterns with their respective demographic histories. Samples from three quilombo remnants populations of the Plaui State, Gaucinha (GAU, n = 14), Mimbi (MIB, n = 15) and Sitão Velho (STV, n = 15) and the urban population of Teresina, Plaui (TES, n = 15), and seven population samples from the HapMap Project (CEU, CHB, JPT, ASW, LWK, M KK, YRI, all with n = 15) were analyzed. More than 250 thousand SNPs (Single Nucleotide Polymorphisms) were genotyped using the GeneChip® Human Mapping 250K Nsp Array I - Affymetrix® in the samples of the four Brazilian populations. Raw data of the HapMap population samples for this array were obtained from the HapMap homepage. Genotypes for all samples were analyzed using GRLMM algorithm. LD analyzes and determination of haplotype blocks were performed using the Haploview software. Considering the number of haplotype blocks detected in each population, a consistent pattern was observed for all autosomes. The European population (CEU) and the two Asian populations (CHB and JPT) of the HapMap showed the highest numbers of blocks, while the lowest numbers were observed in the GAU and MIB quilombos and in the TES population. The African populations, LWK, M KK and YRI, and the African-American ASW exhibited intermediate values. The number of haplotype blocks in STV, presented a number of blocks just smaller than that observed for CEU, CHB and JPT. The great African contribution in the GAU and MIB quilombos may explain their lower LD. On the other hand, the lower LD in TES is probably due to its foundation and that even random proteins have substantial secondary structure. Since these arise evolutionarily and integrate into existing, ancient circuits. A new component can surprisingly be a new gene, raising the question of how such a gene can be used to infer recombination rates in admixed populations using information about recombination events in pedigrees. Recent work (Hinch et al., Wegmann et al.) has demonstrated how knowledge of local ancestry can be used to infer relative recombination rates in admixed populations. We apply similar techniques to admixed Latino American populations to infer a high-resolution map of relative recombination rates. We present a method for inferring relative recombination rates in Latino populations by using signals from the admixture events in their evolutionary history. We used 2001 Latino individuals of Mexican and Puerto Rican ancestry from the GALA study, including 551 trios, for our analysis. Local ancestry inference was performed using software RFMix to infer tracts of African, European and Native American ancestry in the study individuals. The inferred ancestry switch points were then used to perform statistical inference and infer recombination rate variation along the genome. We used the statistical framework of Dirichlet Process Mixture Models to correct for the biases of local ancestry inference and estimate underlying rates of ancestry switches. We validated our method on data simulated from an admixture model of the evolutionary history of Latino populations, with admixtures parameters from literature. Simulations show that our method can accurately resolve ancestry switch points to within a distance of 400 kb with 80% accuracy. Our analysis of the Latino dataset shows evidence of differentiation in recombination rates between populations. We also compare the recombination rates in the different admixed and nonadmixed populations and interpret the results in light of the PRDM9 variants observed in each. References [1] Wegmann, D., Kessner, D. E., Veeramah, K. R., Mathias, R. A., Nicola, D. L., Yanek, L. R., Sun, Y. V., et al. (2011). Recombination rates in the different admixed and nonadmixed populations and interpret the results in light of the PRDM9 variants observed in each. [2] Hinch, A. G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C. D., Chen, G. K., et al. (2011). The landscape of recent human admixture in African Americans. Nature, 476(7365), 170-5. doi:10.1038/nature10336.

1988W
A recombination map of Latino populations inferred using local ancestry. S. Shringarpure, S. Gravel, C. Gignoux, A. Moreno, C. Eng, S. Huntsman, D. Torgerson, E. Burchard, C. Bustamante. 1) Genetics, Stanford University, Stanford, CA, 94305; 2) McGill University, Human Genetics, Montreal, H3A 1B1, Canada; 3) University of California, San Francisco, Bioengineering, San Francisco, CA, 94158. Recombination maps have traditionally been inferred in populations using information about recombination events in pedigrees. Recent work (Hinch et al., Wegmann et al.) has demonstrated how knowledge of local ancestry can be used to infer relative recombination rates in admixed populations. We apply similar techniques to admixed Latino American populations to infer a high-resolution map of relative recombination rates. We present a method for inferring relative recombination rates in Latino populations by using signals from the admixture events in their evolutionary history. We used 2001 Latino individuals of Mexican and Puerto Rican ancestry from the GALA study, including 551 trios, for our analysis. Local ancestry inference was performed using software RFMix to infer tracts of African, European and Native American ancestry in the study individuals. The inferred ancestry switch points were then used to perform statistical inference and infer recombination rate variation along the genome. We used the statistical framework of Dirichlet Process Mixture Models to correct for the biases of local ancestry inference and estimate underlying rates of ancestry switches. We validated our method on data simulated from an admixture model of the evolutionary history of Latino populations, with admixtures parameters from literature. Simulations show that our method can accurately resolve ancestry switch points to within a distance of 400 kb with 80% accuracy. Our analysis of the Latino dataset shows evidence of differentiation in recombination rates between populations. We also compare the recombination rates in the different admixed and nonadmixed populations and interpret the results in light of the PRDM9 variants observed in each. References [1] Wegmann, D., Kessner, D. E., Veeramah, K. R., Mathias, R. A., Nicola, D. L., Yanek, L. R., Sun, Y. V., et al. (2011). Recombination rates in the different admixed and nonadmixed populations and interpret the results in light of the PRDM9 variants observed in each. [2] Hinch, A. G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C. D., Chen, G. K., et al. (2011). The landscape of recent human admixture in African Americans. Nature, 476(7365), 170-5. doi:10.1038/nature10336.

1989T
De novo genes in evolution: Apcdd1, a novel dual Wnt and Bmp inhibitor. V. Luria, S. G. Oh, A. M. Christiano. 1) Systems Biology, Harvard Medical School, Boston, MA; 2) Children’s Hospital Boston, Harvard Medical School, Boston, MA; 3) Genetics & Development, Columbia University Medical Center, New York, NY. The dynamic instability of biological circuits enables them to decide between alternatives. In genetic circuits transducing signals to cells, circuit components can surprisingly be new genes, raising the question of how these arise evolutionarily and integrate into existing, ancient circuits. A new signal transduction component, Apcdd1, encodes a 514 amino-acid membrane-anchored extracellular protein that, intriguingly, has no significant similarity to any protein. Expressed in neurons, germline and ectodermal appendages, it can be induced by Wnts and, when mutated, causes a form of human hair loss. We showed Apcdd1 inhibits Wnt signaling in chick, frog and humans (Shimomura*, Luria* et al., Nature 2010). We now show APCDD1 inhibits Bmp signaling using biochemistry, single-cell imaging and in vivo experiments in chick, frog and zebrafish. The position of APCDD1 at the intersection of Wnt/Bmp pathways may enable it to coordinate pathway activation dynamics. To determine whether Apcdd1 is a de novo gene, we asked when Apcdd1 appeared in evolution, and interrogated the likelihood of gene emergence by coopting random sequences. We found Apcdd1 appeared in Eurometazoa, has no homologues, and that most APCDD1 amino acid sequence is relatively disordered with no recognizable protein domain, suggesting Apcdd1 is a de novo gene. To determine the frequency and content of de novo genes, combining bioinformatics and mathematical modeling, we investigated how de novo genes may arise as a function of gene genome and population parameters, and the protein structure content of de novo proteins. Surprisingly, we find de novo genes may arise more frequently than previously predicted, may be more disordered than ancient proteins, and that even random proteins have substantial secondary structure. Since de novo genes like Apcdd1 can acquire genetic partners and biochemical functions, the next challenge will be understanding what fraction are maintained in genomes that continuously generate and destroy new genes.

1989F
De novo genes in evolution: Apcdd1, a novel dual Wnt and Bmp inhibitor. V. Luria, S. G. Oh, A. M. Christiano.
1990F Detecting evolutionary strata on the human X chromosome in the absence of gametologous Y-linked sequences. M. Wilson Sayres¹, R. Shanker Pandey², R. A. Azad¹, ³ 1) Integrative Biology, University of California, Berkeley, Berkeley, CA; 2) Biological Sciences, University of North Texas, Denton, TX; 3) Mathematica, University of North Texas, Denton, TX.

Mammalian sex chromosomes arose from a pair of homologous autosomes that differentiated into the X and Y chromosomes following a series of recombination suppressions that likely happened due to serial inversions on the Y. The stepwise recombination suppressions from the distal long arm to the distal short arm of the chromosomes are reflected as regions with distinct X-Y divergence, referred to as evolutionary strata on the X. Identifying evolutionary strata is central to understanding the history and dynamics of sex chromosome evolution. All current methods of stratum detection depend on X-Y comparisons but are severely limited by the paucity of X-Y gametologs. We have developed an integrative method that combines a top down, recursive segmentation algorithm with a bottom up, agglomerative clustering algorithm to decipher compositionally distinct regions on the X, which reflect regions of unique X-Y divergence. In application to human X chromosome, our method correctly classified a concatenated set of 35 previously assayed X-linked gene sequences by evolutionary strata. We then extended our analysis, applying this method to the entire sequence of the human X chromosome, in an effort to define specific stratum boundaries. The boundaries of more recently formed strata, 4 and 5, have been defined by previous studies, and are recapitulated with our method. The older strata, 1-3, have remained poorly resolved due to paucity of X-Y gametologs. By analyzing the entire X sequence, our method identified seven evolutionary strata in these ancient regions, where only three could previously be assayed. These new strata are characterized by the presence of X-linked evolution of one or more gambologous Y-linked sequences. Our study also provides information about the contribution of repetitive elements in shaping the recombination landscape of the human X chromosome following the suppression of X-Y recombination.

1991W NANOGP8: Evolution of a Human-Specific Cancer-Promoting Retrogene. J. D. Fairbanks¹, T. H. Ogden¹, A. D. Fairbanks¹, G. J. Parker¹, P. J. Maughan². 1) Department of Biology, Utah Valley Univ, Orem, UT; 2) Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT.

NANOGP8 is a human (Homo sapiens) retrogene, expressed predominantly in cancer cells where its protein product is tumorigenic. It arose through retrotransposition from its parent gene, NANOG, which is expressed predominantly in embryonic stem cells. Based on identification of fixed and polymorphic variants in a genetically diverse set of human NANOG and NANOGP8 sequences, we estimated the evolutionary origin of NANOGP8 at approximately 0.9 to 2.5 million years ago, more recent than previously estimated. We also discovered that NANOGP8 arose from a derived variant allele of NANOG containing a 22-nucleotide pair deletion in the 3' UTR, which has remained polymorphic in modern humans. Evidence from our experiments indicates that NANOGP8 is fixed in modern humans even though its parent allele is polymorphic. The presence of NANOGP8-specific sequences in Neanderthal reads provided definitive evidence that NANOGP8 is also present in the Neanderthal genome. Some variants between the reference sequences of NANOG and NANOGP8 utilized in cancer research to distinguish RT-PCR products are polymorphic within NANOG or NANOGP8 and thus are not universally reliable as distinguishing features. NANOGP8 was inserted in reverse orientation into the LTR region of an SVA retroposon that arose in a human-chimpanzee-gorilla common ancestor after divergence of the orangutan ancestral lineage. Transcription factor binding sites within and beyond this LTR may promote expression of NANOGP8 in cancer cells, although current evidence is circumstantial. The fact that NANOGP8 is a human-specific retro-onsogene may partially explain the higher genetic predisposition for cancer in humans compared with other primates.

1992T Characterizing bias in population genetic inferences from uncertain genotype data. E. Han¹, J. Sinshheimer¹, J. Novembre¹,², ³. 1) Biostatistics, UCLA, Los Angeles, CA; 2) Ecology and Evolutionary Biology, UCLA, Los Angeles, CA.

The site frequency spectrum (SFS) is of primary interest in population genetic studies, because the SFS compresses variation data into a simple summary from which most population genetic inferences can be drawn. However, inferring the SFS from sequencing data is challenging because genotype calls from sequencing data are often inaccurate due to high error rates and if not accounted for, this genotype uncertainty can lead to serious bias in downstream analysis based on the inferred SFS. In our work, we compare two types of approaches to estimate the SFS from sequencing data by detailed simulations: one method that infers individual genotypes from aligned sequencing reads and then estimate the SFS based on the inferred genotypes (two-stage approach) and another method that directly estimates the SFS from aligned sequencing reads by maximizing a likelihood of the SFS (direct estimation approach). We find that the SFS by the direct estimation approach is unbiased even at low coverage, whereas those by the two-stage approach become biased as coverage decreases and most deviations come from the sites with rare variants. Interestingly, the bias by the two-stage approach is in opposite directions depending on the pipeline to infer genotypes: estimating genotypes by pooling individuals in a sample (multi-sample calling) results in underestimation of the number of rare variants, whereas estimating genotypes in each individual and merging them later (single-sample calling) leads to overestimation of rare variants. We characterize the impact of these biases on downstream analyses, such as demographic parameter estimation and rank-based genome-wide selection scans. We observe that at low coverage, the estimated growth rate based on the directly inferred SFS is almost unbiased, whereas using the SFS inferred by a two-stage approach is biased upwards using the single-sample calling and biased downwards using the multi-sample calling. In contrast, bias is less problematic for rank-based genome-wide selection scans because the rank ordering of genomic regions is more robust to genotype errors and biases in the inferred SFS. Our works highlight that depending on the pipeline to infer the SFS from sequencing data one might reach different conclusions in population genetic inferences with the same data set and care is vital for these analyses.


Gene conversion refers to the non-reciprocal transfer of genetic information between two recombining sequences, and there is evidence that this process is biased towards G and C alleles. Biased gene conversion influences both standing genetic variation and the probability of fixation of new mutations. Using high-coverage whole genome sequences of African hunter-gatherers (including 5 West African Pygmy, 5 Hadza, and 5 Sandawe genomes), other global human populations, and primate outgroups we quantify the effects of GC-biased gene conversion on population genetic data. Gene conversion results in modified allele frequency distributions, particularly when the ancestral allele is A or T and the derived allele is G or C. These shifted allele frequency distributions yield modified values of Tajima's D and Fay and Wu's H, potentially leading to false inferences of natural selection and population size changes. These effects vary across the recombination landscape of the human genome, and are more pronounced in high recombination regions. In addition, biased gene conversion can at least partly explain the two-fold difference in mutation rates estimated from human-chimp comparisons and pedigree data. Because molecular clock estimates depend on accurate estimates of mutation rates, incorporating gene conversion leads to more accurate inference of demographic history. Taken together, our findings reveal that molecular genetic phenomena like GC-biased gene conversion have important population genetic implications.
1994W

Characterization and evolution of LAVA elements in gibbons. M.K. Konkel,1 J.A. Walker,1 B. Ulmer2, T.J. Meyer1, A. Damon1, R. Hubley1, A.F.A. Smitt,1 L. Carbone3, M.A. Batzer2 for the Gibbon Genome Sequencing and Analysis Consortium. 1) Dept. of Biological Sciences, Louisiana State University, Baton Rouge, LA; 2) School of Electrical Engineering and Computer Science, Center for Computation and Technology, Louisiana State University, Baton Rouge, LA; 3) Dept. of Behavioral Neuroscience, Oregon Health & Science University, Beaverton, OR; 4) Molecular Biology Center, Babes-Bolyai University, Cluj-Napoca, Cluj, Romania; 5) Institute for Systems Biology, Seattle, WA.

We have recently discovered a novel retrotransposon called ‘LAVA’ in the gibbon lineage. LAVA represents a composite element closely related to the hominoid-specific SVA element. Both share the VA part (VNTR variable number of tandem repeat region) and Alu-like sequence. However, instead of the SVA-specific SINE-R region, LAVA elements contain short tandem repeat sequence sections as well as ancient Alu and L1 sequences. In our investigations of this element, we first analyzed LAVA insertions to confirm that they harbor hallmarks of non-LTR (long terminal repeat) retrotransposons. We conclude that LAVAs most likely rely on the enzymatic machinery of L1 (long interspersed element 1) for their propagation. Next, we investigated the tempo and mode of LAVA expansion in the gibbon lineage leading to Nomascus leucogenys (NLE). Computational reconstruction followed by manual curation of the LAVA subfamily structure revealed 23 distinct full-length subfamilies in the NLE draft genome assembly (Nieuw 1.1). We identified >2500 LAVA elements since the origin of gibbons in the NLE genome. Based on the LAVA subfamily structure, activity of some subfamilies peaked prior to the radiation of gibbons, while other subfamilies show evidence for continued propagation until (at least) very recently. To better understand the expansion dynamics of LAVA elements and to investigate the still unresolved gibbon phylogeny, we selected 200 loci for our phylogenetic PCR analyses. We selected LAVA elements based on their divergence from their respective conspecific sequence. PCR analyses were performed on a primate DNA panel containing 13 gibbon species from all four genera, with four great apes and green monkey as outgroups. Our results confirm computational findings that several of the oldest appearing subfamilies ceased activity prior to the radiation of gibbons; while some subfamilies provide evidence for very recent retrotransposition. Even though mobile elements represent a marker system with advantages such as identity by descent and near-absence of homoplasy, we were not able to determine the radiation times of LAVA elements based on full-length subfamilies.

1996F


Understanding the core set of genes that are necessary for basic developmental functions is one of the central goals in biology. Studies in model organisms identified a significant fraction of essential genes through the analysis of null-mutations that lead to lethality. Recent large-scale next-generation sequencing efforts have provided unprecedented data on genetic variation in human. However, evolutionary and genomic characteristics of human essential genes have never been directly studied on a genome-wide scale.

In this study we used detailed phenotypic resources available for the mouse and deep genomics sequencing data from human populations to characterize patterns of genetic variation and mutational burden in a set of orthologs of known mouse essential genes. Using the 1000 Genomes Phase 1 dataset we compared the genomic characteristics of these genes with all protein-coding genes, as well as a set of 3,811 genes with non-lethal mouse phenotypes. Consistent with the action of strong, purifying selection, the essential genes exhibit significantly reduced levels of sequence variation (Wilcoxon P=3.12×10-35), and increased conservation across the primate and rodent lineages (Wilcoxon P=1.28×10-77). In individual estimates of Fst observed – 12 rare variants within essential genes predicted to be damaged. Consistent with the hypothesis that mutations in essential genes are risk factors for neurodevelopmental disease, we show that de novo variants in patients with Autism Spectrum Disorder (ASD) are significantly more likely to occur in this collection of genes (P=2.7×10-6 by gene enrichment analysis). We included 179 essential genes de novo events exclusively in ASD cases, we show an enrichment of protein connectivity (P=0.0019 based on 1000 DAPPLE permutations).

In currently ongoing work, we are expanding on the role of essential genes in SNA stars and bipolar disorder based on previously reported de novo variants, as well as inherited variants identified in family-based association studies. While incomplete, our set of human orthologs shows characteristics fully consistent with essential function in human and thus provides a resource for rapid curation and interpretation of sequence data in studies of human disease.

1997W

Exploring the Relationship Between Immune System Related Genetic Variants and Complex Traits and Disease Through a Phenome-Wide Association Study (PheWAS). A. Verma1, H. Kuivaniemi2, G. Tromp3, D.J. Carey4, G.S. Gerhard1,2, J.E. Crowe Jr.3, M.D. Ritchie1, S.A. Pendergrass1, 1) Center for Systems Genomics, The Pennsylvania State University, State College, PA; 2) Geisinger Health System, Danville, PA; 3) Vanderbilt University, Nashville, TN.

Exploring the relationship between immune-system related genetic loci and a wide array of traits and outcomes provides a way to elucidate inter-relationships among the immune system and diagnoses as well as identify new pleiotropic loci. To explore these connections further and identify novel associations and pleiotropy, we selected 132,467 single nucleotide variants (SNPs) with allele frequency > 0.01 that were previously associated with autoimmune diseases from the Geno2Pheno database and performed a Phenome-Wide Association Study (PheWAS). We calculated associations between the SNPs and 480 clinical diagnoses. To define case-control status we used ICD9 diagnosis codes from 3,035 subjects using de-identified electronic medical records from the Geisinger MyCode biorepository. We required ≥ 10 case subjects for ICD9 code inclusion, and used logistic regression for all associations, adjusting models for age and sex. With an exploratory P-value cutoff < 0.001 we found a total of 7,910 SNP-diagnosis associations, and 420 SNPs associated with more than one diagnosis. The most significant novel association was the GLRB SNP rs17035787 and the diagnosis of hemorrhagic disease due to circulating anticoagulants (P = 4.78×10-9; 29 cases, 3,007 controls). Additional novel associations included CAMTA1 SNP rs4722110 and chronic sinusitis (P = 3.5x10-5; 93 cases, 4,017 controls) and the SNP rs11842088 and ‘syncopal collapse’ (P = 4.29×10-8; 10 cases, 2,930 controls). We found replication of previously reported GWAS results, including the second most significant result of this study, the F5 SNP rs6025 associated with the diagnosis of ‘venous thrombosis’ (P = 5.17×10-8; 55 cases, 2,981 controls). Potential pleiotropy was identified, such as the SNP rs79268593 associated with three diagnoses: ‘chest swelling, mass, or lump’; ‘malignant neoplasm of the bladder’; and ‘hypertensive mass, or lump’. Potential pleiotropy was identified, such as the SNP rs79268593 associated with three diagnoses: ‘chest swelling, mass, or lump’; ‘malignant neoplasm of the bladder’; and ‘hypertensive mass, or lump’. Potential pleiotropy was identified, such as the SNP rs79268593 associated with three diagnoses: ‘chest swelling, mass, or lump’; ‘malignant neoplasm of the bladder’; and ‘hypertensive mass, or lump’. Potential pleiotropy was identified, such as the SNP rs79268593 associated with three diagnoses: ‘chest swelling, mass, or lump’; ‘malignant neoplasm of the bladder’; and ‘hypertensive mass, or lump’.
H.Norton, E. Correa. Department of Anthropology, University of Cincinnati, Cincinnati, OH.

Here we report on the frequency and distribution of an allele in the TYRP1 gene associated with blond hair color in several populations from across western Island Melanesia. This mutation and its association with hair color were originally discovered by Kenney et al. (2013) in populations from the Solomon Islands, where the blond-hair associated allele occurs at an average frequency of 0.26. However, knowing that the blondism phenotype in Melanesia extends further to the east, we typed this mutation in 515 individuals from several different islands throughout the Bismarck Archipelago. The mean frequency of the blondism allele in the region is much lower (0.125) than observed in Solomon Island populations, although it varies dramatically from island to island, ranging from a frequency of 0.015 on the island of New Britain to a frequency of 0.260 on the island of New Ireland. Linear regression using a recessive model was used to test for an association between genotype at this SNP and quantitatively assessed skin and hair pigmentation (measured as the M index). Associations were tested for in the total sample as well as on three islands separately (New Britain, n = 235; New Ireland, n = 152; New Hanover, n = 88). Using age and island as covariates, we demonstrate that genotype at the R932C SNP can explain 8.8% of the variance in hair color (p < 0.0001) in the full dataset. A significant association between the R932C genotype and hair M index was only observed on the island of New Ireland (p < 0.01), but this may be due in part to reduced power to detect these associations in the smaller island subsamples. We do not observe a significant association between skin pigmentation and R932C genotype in the full dataset or the three islands individually, confirming previous observations (Kenny et al., 2013) that this allele affects hair pigmentation independent of skin pigmentation. The geographic heterogeneity of the blondism allele in western Island Melanesia is consistent with extensive population substructure in the region.

We extend the Ewens Sampling Formula (ESF) under the infinite-alleles model to accommodate inbreeding. Our method uses the ESF to determine the likelihood for diploid samples comprising data from patches throughout the genome, accounting explicitly for multi-locus identity disequilibrium generated by inbreeding. Each locus draws a mutation rate from a common mutation rate distribution, which is itself estimated from the data. We construct an MCMC algorithm and apply it to the analysis of androdioecious, gynodioecious, and pure hermaphroditic populations, jointly inferring the frequencies of each reproductive class and rates of inbreeding and locus-specific mutation.

B.D. Redelings, M.K. Uyenoyama. Biology Dept., Duke University, Durham, NC.

We present a novel Bayesian method for inferring the inbreeding rate and locus-specific mutation rates for populations reproducing by partial selfing. We extend the Ewens Sampling Formula (ESF) under the infinite-alleles model to accommodate inbreeding. Our method uses the ESF to determine the likelihood for diploid samples comprising data from patches throughout the genome, accounting explicitly for multi-locus identity disequilibrium generated by inbreeding. Each locus draws a mutation rate from a common mutation rate distribution, which is itself estimated from the data. We construct an MCMC algorithm and apply it to the analysis of androdioecious, gynodioecious, and pure hermaphroditic populations, jointly inferring the frequencies of each reproductive class and rates of inbreeding and locus-specific mutation.

1998T

We present a novel Bayesian method for inferring the inbreeding rate and locus-specific mutation rates for populations reproducing by partial selfing. We extend the Ewens Sampling Formula (ESF) under the infinite-alleles model to accommodate inbreeding. Our method uses the ESF to determine the likelihood for diploid samples comprising data from patches throughout the genome, accounting explicitly for multi-locus identity disequilibrium generated by inbreeding. Each locus draws a mutation rate from a common mutation rate distribution, which is itself estimated from the data. We construct an MCMC algorithm and apply it to the analysis of androdioecious, gynodioecious, and pure hermaphroditic populations, jointly inferring the frequencies of each reproductive class and rates of inbreeding and locus-specific mutation.

1999F
Geographic structure of an allele associated with blond hair color in western Island Melanesia. H. Norton, E. Correa. Department of Anthropology, University of Cincinnati, Cincinnati, OH.

Here we report on the frequency and distribution of an allele in the TYRP1 gene associated with blond hair color in several populations from across western Island Melanesia. This mutation and its association with hair color were originally discovered by Kenney et al. (2013) in populations from the Solomon Islands, where the blond-hair associated allele occurs at an average frequency of 0.26. However, knowing that the blondism phenotype in Melanesia extends further to the east, we typed this mutation in 515 individuals from several different islands throughout the Bismarck Archipelago. The mean frequency of the blondism allele in the region is much lower (0.125) than observed in Solomon Island populations, although it varies dramatically from island to island, ranging from a frequency of 0.015 on the island of New Britain to a frequency of 0.260 on the island of New Ireland. Linear regression using a recessive model was used to test for an association between genotype at this SNP and quantitatively assessed skin and hair pigmentation (measured as the M index). Associations were tested for in the total sample as well as on three islands separately (New Britain, n = 235; New Ireland, n = 152; New Hanover, n = 88). Using age and island as covariates, we demonstrate that genotype at the R932C SNP can explain 8.8% of the variance in hair color (p < 0.0001) in the full dataset. A significant association between the R932C genotype and hair M index was only observed on the island of New Ireland (p < 0.01), but this may be due in part to reduced power to detect these associations in the smaller island subsamples. We do not observe a significant association between skin pigmentation and R932C genotype in the full dataset or the three islands individually, confirming previous observations (Kenny et al., 2013) that this allele affects hair pigmentation independent of skin pigmentation. The geographic heterogeneity of the blondism allele in western Island Melanesia is consistent with extensive population substructure in the region.

2000W
Test of synergistic interactions among deleterious alleles in humans. M. Sohail1, A. Kondrashov2, P.I.W de Bakker2, S. Sunyaev4, GoNL Consortium. 1) Systems Biology PhD Program, Harvard University, Boston, MA; 2) University of Michigan, Ann Arbor, MI; 3) University Medical Center Utrecht, Utrecht, Netherlands; 4) Brigham and Women’s Hospital, Boston, MA.

Understanding the nature of epistatic interactions is crucial to obtaining a complete picture of genetic variation underlying phenotypic diversity. The role of epistatic interactions in the genetic architecture of human complex phenotypes has been widely debated. In evolutionary genetics, epistatic selection may be an important force shaping DNA sequence variation with implications for the adaptive value of sex and reproduction. We developed a new statistical method to detect epistasis in sequencing data, based on testing for variance depletion in the distribution of deleterious mutations. Our statistical test is suitable for performing phenotype-focused studies in medical genetics and addressing key questions in evolutionary theory. We applied our test to detect if synergistic epistasis is the predominant mode of genetic interaction with respect to fitness in human populations. Synergistic interactions mean that multiple mutations have a larger cost on fitness than expected from their individual effects. The mutational deterministic hypothesis for the maintenance of sex postulates that sex is an adaptation to purge the genome of deleterious mutations, and requires synergistic epistasis. We had access to a unique sequencing dataset for this study. The Genome of the Netherlands (GoNL) is an effort to characterize genomic variation in the Dutch population through whole-genome sequencing of 250 families (231 trios, 19 twin quartets) at 12x using Illumina HiSeq. In the GoNL data, we observed that spouses are correlated in the number of private variants for all functional classes of variants, but that the variation of this measure can be primarily attributed to the province of origin, suggesting subtle population substructure within the Netherlands. Indeed, we detected a north-south gradient of increasing heterozygosity due to private variants. After correcting for this structure, our results indicate that the variance in the number of private variants in a specific category decreases with the potential functional importance of the allelic class. However, in the current dataset the variance depletion is not significant compared to the theoretical expectation. This may be due to additional sources of variance inflation, or the predominant synergistic epistasis that we detect. We complement our analysis with simulations and analytical approaches to characterize the strength of the variance depletion due to synergistic epistasis as a function of selection.
2001T Genetic Diversity is a Predictor of Survival in Humans. N.A. Bihlmeyer, A. Scaria, M. Nalls, M. Garcia, K.L. Lunetta, J.M. Murabito, D.R. Weir, J.A. Smith, M. All讳herr, I. Yu, D.A. Bennett, S.S. Mirza, N. Direck, A. Teumer, G. Homuth, A.V. Smith, V. Gudnason, T. Lumley, D.E. Arking on behalf of the CHARGE Aging & Longevity Working Group. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Statistics, University of Auckland, Auckland, NZ; 4) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 5) Laboratory of Epidemiology, Demography and Biometry, National Institute of Aging, National Institutes of Health, Bethesda, MD, USA; 6) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 7) The National Heart Lung and Blood Institute’s Framingham Heart Study, Framingham, MA; 8) Survey Research Center, Institute for Social Research, University of Michigan 426 Thompson Street, 4123 MSQS, Ann Arbor, MI; 9) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI; 10) Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, UK; 11) Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago IL, USA; 12) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 13) Interfaculty Institute for Genomics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany; 14) Icelandic Heart Association, Kopavogur, Iceland; 15) University of Iceland, Reykjavik, Iceland.

With the advent of genome-wide association studies (GWAS), and more recently whole-exome and whole-genome sequencing, remarkable progress has been made in elucidating the genetics of complex traits, with numerous genetic variants each explaining a small fraction of the variance. The presence of numerous segregating small effect alleles that influence traits that directly impact health raises the question as to whether global measures of genetic diversity across many organisms, and is often referred to as Positive Selection. Whole-genome sequencing, and is responsible for lactase persistence. The widespread occurrence of this allele on a very extended haplotype suggests strong directional selection in the last 5000 years. In Africa and the Middle East, the situation is more complicated and at least three other alleles (-13907*G, -13915*G, -13910*T) in the same locus can cause continued lactase production. There also exists a large diversity of lactase persistence alleles in Ethiopia; signature of a soft selective sweep. Additionally, we focus on variants showing high frequency differences between populations, and ask whether specific functionally equivalent alleles could potentially hide causal gene regions.


The persistent expression of lactase into adulthood in humans is a recent genetic adaptation that allows the consumption of milk from other mammals after weaning. In Europe, a single allele (-13910*T, rs4988235), located in an upstream enhancer of the lactase gene, LCT, increases lactase expression in promoters constructs in vitro and is responsible for lactase persistence. The widespread occurrence of this allele on a very extended haplotype suggests strong directional selection in the last 5000 years. In Africa and the Middle East, the situation is more complicated and at least three other alleles (-13907*G, -13915*G, -13910*T) in the same LCT enhancer region can cause continued lactase expression. We have now examined the LCT enhancer sequence in a lactose tolerance tested Ethiopian cohort of more than 350 individuals. We show that a further SNP, -14009>T>G is significantly associated with lactose digester status and in vitro functional tests confirm that the 14009>G allele also increases expression of an LCT promoter construct. Other rarer SNPs are either more frequent in non-digesters or not frequent enough to assess for association. These observations are of relevance to genome wide disease association studies since such panel sets of functionally equivalent alleles could potentially hide causal gene regions.

2004T Characterization of private and highly diverged variants among human populations from the analysis of exome sequencing. L.R. Boligård, D. Bobo, D. Twigg, C.D. Bustamante, J.M. Kidd, B.M. Henn. 1) Ecology and Evolution, Stony Brook, Stony Brook, NY; 2) Human Genetics, University of Michigan, Ann Arbor, MI; 3) Genetics, Stanford University, Stanford, CA; 4) Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI.

Previous analysis of human genomic variation revealed an excess of private variants present at very low frequencies in Eurasian populations, and that variants in Europeans are enriched for possibly damaging functional effects. These patterns have been related to the Out-of-Africa bottleneck and the exponential growth of European populations over the last 200 generations. However, a recent analogous study found no significant differences between African-American and European populations regarding the burden of potentially deleterious mutations. Here, we analyze exome data from ten populations with different ancient and recent demographic histories (sequences include publically available 1000 Genomes Project data and samples from the Human Genome Diversity Panel). By comparing populations with and without explosive growth or severe bottlenecks we are able to assess the effect of demography and zygosity on the distribution of potentially damaging alleles. Additionally, we focus on variants showing high frequency differences between populations, and ask whether specific functional categories are driving differentiation among human populations, with the ultimate goal of understanding the relative importance of demography and natural selection as drivers of population divergence.
Identification of known genes in a novel gene discovery project: Experience of the FORGE Canada Consortium. S.L. Sawyer1, C.L. Beauleau1, D.E. Bulman1, F.P. Bernier2, K.M. Boycott1. FORGE Canada Consortium. 1) Medical Genetics, Children's hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 2) Medical Genetics, Alberta Children's Hospital Research Institute, Calgary, Alberta, Canada.

The primary objective of the FORGE (Finding Of Rare disease GEnes) Canada project was the identification of novel disease-causing genes. Over 200 rare diseases with pediatric onset were studied over a 2-year period using whole exome sequencing (WES) and one of three approaches: (1) cohorts of patients with the same rare disease; (2) consanguineous or dominant families with mapping data; and, (3) affected sib-pairs born to unrelated parents. In approximately 1/3 of the 200 FORGE projects analyzed, known genes previously associated with disease were identified as causative for the phenotype despite careful selection of projects prior to WES to enrich for those more likely to identify a novel disease gene. Review of the characteristics of the approximately 60 projects that identified known disease genes highlighted several factors that likely influenced this outcome including; (1) overall significant rarity of the syndrome (ultrarare); (2) disease presentations with high levels of genetic heterogeneity; (3) limited access to exclusionary genetic testing prior to enrollment in the study; and, (4) atypical clinical presentations. Further stratification of our results based on types of clinical presentation indicated that the likelihood of identifying a mutation in a known disease gene in a patient presenting with cerebellar ataxia, for example, was approximately 40%, while the chance of making a molecular diagnosis in a patient with neuropathy was significantly lower. Autosomal recessive inheritance of the disease was the most significant factor positively influencing success in these projects. Given that half of the 120 projects solved through the FORGE Canada initiative were known disease genes we have gained significant insight into the clinical utility of WES and potentially the depth of rare disease.

## Application on OMIM disease genes

<table>
<thead>
<tr>
<th>O M I M disease genes</th>
<th>O M I M recessive</th>
<th>O M I M Haploinsufficiency</th>
<th>O M I M de novo</th>
<th>O M I M Haploinsufficiency and de novo</th>
<th>Essential Gene List (Georgi et al. 2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of genes</td>
<td>2131</td>
<td>817</td>
<td>175</td>
<td>467</td>
<td>108</td>
</tr>
<tr>
<td>Intolerance Score</td>
<td>0.58</td>
<td>0.55</td>
<td>0.73</td>
<td>0.65</td>
<td>0.78</td>
</tr>
<tr>
<td>[0.57 - 0.65]</td>
<td>[0.53 - 0.67]</td>
<td>[0.70 - 0.74]</td>
<td>[0.66 - 0.72]</td>
<td>[0.74 - 0.80]</td>
<td>[0.65 - 0.85]</td>
</tr>
</tbody>
</table>

### 2006W

Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. S. Petrovski1,2, Q. Wang1, E.L. Heinzen1, A.S. Allen1, D.B. Goldstein1. 1) Center for Human Genome Variation, School of Medicine, Duke University, Durham, NC; 2) Departments of Medicine, Autism Health and Royal Melbourne Hospital, University of Melbourne, Australia; 3) Department of Bioinformatics and Bioinformatics, Duke University, Durham, NC.

Using empirical polymorphism data from the NHLBI Exome Sequencing Project we introduce a genome-wide scoring system that ranks human genes in terms of their intolerances to functional genetic variation in the human population. We name this score the Residual Variation Intolerance Score (RVIS). It is often inferred that genes carrying relatively fewer common functional variants in healthy individuals may be judged more likely to cause certain kinds of disease. We show that this ‘intolerance score’ correlates remarkably with genes already known to cause Mendelian diseases (P=10^-27). Equally striking, however, are the differences in the relationship between standing genetic variation and disease causing genes for different disease types. Considering disorder classes defined by Goh et al (2007) human disease network, we show a nearly opposite pattern for genes linked to developmental disorders and those linked to immunological disorders, with the former being preferentially caused by genes that do not tolerate functional variation and the latter caused by genes with an excess of common functional variation. We conclude by showing that use of an exome-wide intolerance ranking framework can facilitate interpreting personal genomes and can facilitate identifying high impact mutations through the gene in which they occur.

<table>
<thead>
<tr>
<th>Score 0.29 -0.16 -0.71 -0.57 -0.77 -0.63</th>
</tr>
</thead>
<tbody>
<tr>
<td>[AUC 0.57 - 0.62] [CI 0.53 - 0.69]</td>
</tr>
</tbody>
</table>

### Table

<table>
<thead>
<tr>
<th>Number of genes</th>
<th>2131</th>
<th>817</th>
<th>175</th>
<th>467</th>
<th>108</th>
<th>2288</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intolerance Score</td>
<td>0.58</td>
<td>0.55</td>
<td>0.73</td>
<td>0.65</td>
<td>0.78</td>
<td>0.66</td>
</tr>
<tr>
<td>[0.57 - 0.65]</td>
<td>[0.53 - 0.67]</td>
<td>[0.70 - 0.74]</td>
<td>[0.66 - 0.72]</td>
<td>[0.74 - 0.80]</td>
<td>[0.65 - 0.85]</td>
<td>[0.70 - 0.90]</td>
</tr>
</tbody>
</table>

### Logistic Regression

<table>
<thead>
<tr>
<th>Score (RVIS) AUC C</th>
<th>[95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0.57 - 0.65]</td>
<td>[0.70 - 0.74]</td>
</tr>
</tbody>
</table>
I. Balcells1, M. Mele2, T. Marques-Bonet1,3, Y. Espinosa-Parilla1. 1) Institut de Biologia Evolutiva (UPF-CSIC), Barcelona, Spain; 2) Centre for Genomic Regulation (CRG), Barcelona, Spain; 3) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

MicroRNAs (miRNAs) are important regulators of gene expression at post-transcriptional level. Although lineage-specific miRNAs may contribute to shape primate evolution; how it happens and what are the specific miRNA changes that contribute to this process is not well understood yet. Using genomic data, we have compared the 1,595 precursor miRNAs (pre-miRNAs) described in humans (miRBase v19) in several Great Ape populations including humans (n=9), bonobos (n=13), chimpanzees (n=25), gorillas (n=27) and orangutans (n=14). Our data showed that a total of 1,782 single nucleotide variants (SNV) occurred in pre-miRNA; 590 in the mature miRNAs and 198 in the miRNA seed region. Surprisingly, although it has been described that miRNA seeds are highly constrained, in this study, SNV densities (SNVD) do not differ between the seed region (SNVD= 0.023), the mature miRNA (SNVD=0.022) and all pre-miRNA (SNVD=0.024). And, interestingly, the distribution of nucleotide changes was seen not to be uniform. They are accumulated in 734 (65.4%) pre-miRNAs, 429 (28.9%) mature miRNAs and 170 (11.9%) seed miRNAs. In the branch going from the chimpanzee-human ancestor to humans, 23 substitutions have been fixed in the seed region of miRNAs. These changes could affect directly miRNA function by modifying the number and spectrum of target genes. Some of these nucleotide changes are predicted to provoke important gain or losses of human miRNA targets according to TargetScan prediction. Nevertheless, expression patterns for these miRNAs and their target genes should be further analyzed to decipher their function. On the other hand, 84 nucleotide changes affected mature miRNAs in the human lineage. Out of them, twenty are expressed in specific regions of the human brain. These nucleotide changes could affect miRNA expression levels and thus, the fine-tuning of brain molecular regulatory networks could be altered. Overall, our work describes that many nucleotide changes have been accumulated during Great Ape evolution in different functional regions of miRNAs. However, the great challenge now is to understand which are the functional implications of these changes.

2008F

Mitochondrial haplogroup B and oxidative stress associations with antiretroviral-associated peripheral neuropathy in Thai individuals. R. Levinson1, T. Hultgren2, M. Gerochsonen2, N. Phanuphak2,9, J. Ananworanich3,4,5,6, A. Baker2, V. Valcour2, D.E. Li Buffy4, T. Jadwattanakul5, D. Murdock1, J. McArthur6, C. Shikuma4, D.C. Samuels2, the SEARCH 003 Study Team. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) VANDERBILT UNIVERSITY, Nashville, TN; 3) South East Asia Research Collaboration with Hawaii, Bangkok, Thailand; 4) University of Hawaii, Honolulu, Hawaii, USA; 5) Thai Red Cross AIDS Research Centre, Bangkok, Thailand; 6) HIV-NAT, Bangkok, Thailand; 7) University of California-San Francisco, San Francisco, CA; 8) Queen Savang Vadhana Memorial Hospital, Chonburi, Thailand; 9) The Johns Hopkins University, Baltimore, MD.

SEARCH 003 was a randomized clinical trial in Thailand designed to evaluate mitochondrial (mt) toxicity of first-line antiretroviral (ARV) regimens in HIV-infected persons. mtDNA sequencing from PBMC was performed on 149 subjects with the Affymetrix GeneChip Human Mitochondrial Resequencing Array 2.0. The SEARCH 003 population fell into 3 mitochondrial haplogroup clades, B (15%), M (43%), and F (21%) determined by PhyloTree; the remainder were categorized as ‘other’. Participants had epidermal nerve fiber density (ENFD) assessed by distal leg skin biopsy at weeks 0 and 24 as a measure of peripheral neuropathy; an ARV-associated mt toxicity. mtDNA levels were measured by RT-PCR. MtDNA oxidative damage was measured by the gene specific repair assay for 8-oxo-deoxyguanine (8-oxo-dG). Analyses included Wilcoxon rank-sum test and logistic regression. Baseline mtDNA levels, ENFD, and treatment arm were not significantly different by mtDNA haplogroup. Haplogroup B individuals had an increase in median ENFD over 24 weeks compared to non-B individuals (p=0.02). Median 24-week PBMC 8-oxo-dG change also increased in haplogroup B individuals compared to non-B individuals (p=0.02). In a logistic regression model, haplogroup B was significantly associated with an increase in ENFD (OR= 24.3, 95%CI = [4.95, 168]). In this model, other significant covariates were baseline CD4 count, age, and ENFD. These data show that haplogroup B individuals gained ENFD over the first 24 weeks of ARV treatment. However, for individuals with ENFD data the frequency of peripheral neuropathy was significantly higher in the haplogroup B individuals (p=0.04), though the number of cases was small (N=6). In conjunction with increased mt oxidative damage in haplogroup B, this leads us to hypothesize that the gain in ENFD may be a compensatory response to mt damage. Follow up studies are needed to validate these associations and elucidate mechanisms by which mtDNA variation influences response to therapy. Most haplogroup B individuals in our study were subgroup B5, though any role of subgroup in these responses remains unclear. Haplogroups are closely related to continental origin and their global distribution follows historical human migrations. Haplogroup B is found in individuals of East Asian descent and Native American lineages, including in about 15% of Hispanics. Follow up studies in Southeast-Asian and Hispanic HIV/AIDS populations would be worthwhile.
2009W
Defects in the autophagy pathway contribute to glaucoma caused by mutant myocilin accumulation. X. Lin1,2, G. Zode1,2, C.C. Searby1,2, V.C. Sheffield1,2, 1) Pediatrics, The University of Iowa, Iowa City, IA; 2) Howard Hughes Medical Institute, MD.

Autophagy is a cellular degradation pathway that involves the delivery of unnecessary or dysfunctional cellular components to the lysosome. Autophagy is often upregulated in circumstances that result in protein aggregate accumulation. Autophagy plays a role in the clearance of aggregate-prone mutant proteins associated with several neurodegenerative diseases in which accumulation of unfolded protein is the major characteristic. Early reports demonstrated that autophagy contributed to the pathogenesis of Alzheimer disease, Parkinson disease and Huntington disease. However, little is known about autophagy in glaucoma, including glaucoma caused by myocilin mutations. It is well documented that mutant myocilin accumulates in the ER of the trabecular meshwork and leads to an increase in intraocular pressure, which is associated with retinal ganglion cell degeneration. We have investigated the role of autophagy in clearing mutant myocilin aggregates. We generated stable HEK 293T cell lines expressing mutant and wild-type myocilin. We confirmed that mutant myocilin accumulates and leads to increased ER stress, while wild type myocilin does not accumulate or result in ER stress. In addition, mutant myocilin leads to increased LC3-I levels compared to wild type myocilin and control cells. We also observed increased levels of p-Akt, beclin-1, (responsible for vesicle elongation), and Atg-4 (pivotal to autophagosome membrane formation) in mutant myocilin cells, while no change was observed in p-mTOR (negative regulator of autophagy), Atg 5 (involved in autophagosome formation), or Atg 3 (involved in LC3-I conjugation). When treated with chloroquine, a lysosome basifying agent which blocks the fusion of autophagosomes with lysosomes, we observed exacerbation of myocilin accumulation only in mutant myocilin cells, but an increase in LC3-II/LC3I ratios only in control cells. Increased LC3-I levels in mutant myocilin cells compared to cells expressing WT myocilin cells suggest that autophagosomes could not form properly. These data indicate that autophagy plays a role in the clearance of mutant myocilin, but is ineffective.

2010T
Modeling Mutation Events for Complex Variants in Human Genome. M. Bhuyan, I. Pe’er. Columbia University, New York City, NY.

Background: Availability of whole genome sequences paves the way to observing sites that harbor complex, multi-allelic variation. Such sites may be considered as candidates for genetic association studies. By definition, complex variants (CVs) require the occurrence of multiple mutation events, often including indels and repeats besides single nucleotide substitutions. In this study we focus on cataloging and categorizing CVs, modeling implied mutation events, and reconstructing ancestry among alleles. Methods: We use variant information from a cohort of 69 individuals from 12 different ethnicities made publicly available by Complete Genomics. Specific quality control steps were taken to ensure the vast majority of observed complex variants are genuine. We classify CVs into categories guided by likely types of mutation events. We define a probabilistic model for ancestral mutation events, assigning each type a rate of occurrence. We estimate the parameters of this model along with local trees that describe the history of all the alleles associated with corresponding CVs. This is achieved by iteratively finding a minimum weight local trees and updating the model from the mutation events across edges of all local trees. The roots are rooted by alignment to the chimpanzee sequence, aided by an efficient ad hoc procedure to resolve repeat boundaries. Results: The observed frequency of CVs is considerably higher than expected by chance. We identify several common categories of CVs, each providing evidence for particular types of mutation events. (i) Simple repeat regions with variable count of their repeat unit make up most CVs. These include homopolymers (~36% of CVs) as well as microsatellites (~6%). (ii) Short CVs, all of whose aligned alleles are ≤3bp in length. These include tri/tetra-allelic SNPs (~7%) di/tri-nucleotide polymorphisms (~24%). (iii) Block polymorphisms (~1%) can be expressed as combinations of two or more SNPs, usually linked. Mutation events in our model include, in addition to point substitutions and point indels, specific types for period indels, block indels, and more specifically, block indels at the end of the allele sequence. Mutation rates depend on context of mutated and flanking single nucleotides, and are fit to the data as a matrix.

2011F

An emerging, major challenge is the identification of pathogenic, rare mutations contributing to complex disease, as this class of alleles offers great potential to accelerate our understanding of mechanisms of the disease etiology. Despite the increasing evidence of the role of noncoding variation in disease, few computational approaches are designed to aid the interpretation of mutations in this large fraction of the genome. New computational approaches design to understand the biological features that govern where mutations occur, how many occur, and how they persist over time (i.e., selection) in human populations at broad and fine scales will be essential to the disease discovery and interpretation process. To address this shortcoming, we propose a statistical framework to model and assess the variability in the population mutation rate at broad and fine scales. The generalized statistical model is Bayesian in nature, and uses repeated sampling from standard distributions to predict the mutation spectrum and also assign confidence interval to different types of mutations, in any number of individuals from different populations. Based on fundamental population genetics quantities, we provide a framework to hypothesis test and incorporate new biological features and annotations like DNase hypersensitivity, differential selection, epigenetic marks etc. that have been measured empirically or speculated. As a proof-of-concept, using data from Phase I of the 1000 Genomes Project, we demonstrate that sequencing context is a major feature that explains the variability in the population mutation rate at all genomic scales (P<2.2E-308), and confirm expected population-size reductions on the X chromosome (~7.3E-182) and in comparisons of between African and non-African population (P<5.7E-204). Furthermore, we show that our analysis framework facilitates discovery of previously unknown features influencing the spectrum of observed genetic variation. Our findings suggest that this basic framework can be used to model and predict the genetic variant spectrum in any region of the genome, information which is critical to building appropriately calibrated rare variant association tests in the non-coding (and coding) genome for complex disease.

2012W
Neanderthal Introggression at Chromosome 3p21.31 was Under Positive Natural Selection in East Asians. O. Ding1, Y. Hu1, S. Xu2, J. Wang3, L. Jin1,2. 1) State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 2) CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.

Studies on Neanderthal and Denisovan genomes demonstrate archaic hominin introgression in modern Eurasians. In the current study, we present evidence of Neanderthal introgression at 3p21.31 region. Frequency of introgressed haplotypes was high in East Asians, ranging from 49.4% to 66.5%, and was low in Europeans. Signal of strong positive selection was detected in this region only in East Asians. Expansion of the introgressed haplotypes started at 37.4 KYA, suggesting the starting point of selection. We showed that rs12488302-T or its associated alleles might be the candidate targets of selection, among which four are non-synonymous, including rs35455589-G of in HYAL2, a gene related to cellular response to UV-B. Furthermore, suggestive evidence supports latitude-dependent selection, implicating the role of UV-B. Interestingly, the distribution of rs35455589-G suggests that this allele was lost during the exodus of modern human from Africa, and reintroduced to Eurasians from Neanderthals.
2013T Homozygous losses of human-specific neural SRGAP2C gene discovered in patients with intellectual disability. X. Nuttles1, K. Witherspoon1, C. Baker1, B. Cos1, M. Fichera1, J. Schuurs-Hoeijmakers4, R. Bernier4, J. Gecc5, B.B.A. de Vries6, C. Romano6, J. Shendure7, E. Eichler8, 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; 2) Regional Center for Genetic Rare Diseases with Intellectual Disability or Brain Aging, IRCOS Associazione Oasi Maria Santissima, Via Conte Ruggero, 73, 94018 Troina, Italy; 3) Medical Genetics, University of Catania, Catania, Italy; 4) Department of Human Genetics, Radboud University, Nijmegen, Netherlands; 5) Department of Psychiatry, University of Washington, Seattle, WA 98195, USA; 6) Department of Paediatrics, The University of Adelaide, Adelaide, Australia; 7) Genetics and Molecular Pathology, and South Australian Pathology at Women’s and Children’s Hospital, Adelaide, Australia; 8) Howard Hughes Medical Institute, University of Washington School of Medicine, Seattle, WA 98195, USA.

Incomplete duplications of SRGAP2, a gene involved in regulating neuronal migration and development of dendritic spines, are hypothesized to have played an important role in human brain evolution. Several observations support this hypothesis, including the evolutionary emergence of the duplicate genes at the onset of Homo encephalization (~2-3 million years ago), fixation of one duplicate (SRGAP2C) in the human population, its expression in fetal brain, and striking effects on neuronal development when introduced into mice. Nevertheless, no direct genetic evidence in humans has yet linked disruptive mutations of these human-specific paralogs with neurodevelopmental disease phenotypes. We recently developed a novel high-throughput method using molecular inversion probes to accurately genotype duplicated genes previously intractable to large-scale genetic analysis, such as SRGAP2. We surveyed copy number and sequence variation in SRGAP2 genes in a diversity panel of 1,056 humans and discovered a single base deletion in exon 3. All five individuals are homozygous, likely due to consanguinity, and are predicted to have completely lost the human-specific SRGAP2C. In contrast, we never observed homozygous loss of SRGAP2C in controls. These results provide further support for the importance of this new human gene for the evolution of unique aspects of human brain development.

2014F Most Autosomal Recessive Diseases Have No Frequent Mutations that Reflect Increased Carrier Fitness or Mutation Hotspots. M.A.H. Gener, R.V. Lebo. Pathology, Akron Children’s Hospital, Akron, Oh.

Heterozygous advantage results when the carrier of a mutant gene has an increased reproductive fitness over a noncarrier. Then the relative frequency of initially occurring mutant alleles increase more rapidly in subsequent generations than the frequency of subsequently mutated alleles. As a consequence, increased frequencies of specific mutant alleles have been reported in diseases exhibiting increased carrier fitness as well as in a disease gene with mutation hotspots. Carriers of one cystic fibrosis (CFTR) mutant allele are reported to exhibit increased activity, whereas carriers of other disease genes like cholera. Carriers of the sickle cell anemia allele and related α- and β-thalassemia alleles survive malarial infection more readily in regions cultivating rice. This selective advantage explains why the ΔF508 mutation comprises ~70% of cystic fibrosis alleles and the sickle cell mutation (~90% of sickle cell alleles). Alternatively, spontaneous mutations in the MECP2 Rett syndrome gene occur at mutation hotspots related to specific de novo, lethal gene sequences. The 34 listed autosomal recessive disorders with frequencies exceeding 1 in 100,000 worldwide patients were identified through GeneReviews and further studied through the Human Gene Mutation Database and Online Mendelian Inheritance in Man websites. No evidence of substantial selection was found for any of these disorders that had between 34 mutations. Kyprikoischolos. Ehlers-Danlos Syndrome and 1009 mutations in Leber Congenital Amaurosis. In contrast, hemochromatosis reported in the large northern European Caucasian regional population results from the Cys282Ty mutation in ~90% of known alleles, possibly reflecting an as yet unreported selective carrier survival advantage. Furthermore, allele selection could readily explain the ~10% of all individuals who carry the 5T CFTR allele which decreases CFTR gene expression. 5T/5T alleles are found in patients with CBX. The lack of common alleles in most diseases makes analysis of multiple disease genes in the worldwide population more complex. In conclusion, inherited heterozygote carrier frequencies in very large patient populations typically reflect the length of time the mutation existed since its occurrence in the human population and whether a founder allele has been selected. Very frequent single mutations are not anticipated in the absence of heterozygous advantage or mutation hotspots.

2015W Multiple LD-independent signals of extreme sub-population variation at a region associated with type-2 diabetes suggests a non-neutral evolutionary history. P.L. Bab1, B.F. Voight1, 2) 1) Pharmacology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

Type-2 diabetes (T2D) is a complex metabolic disease affecting over 280 million people worldwide. Although >70 loci have been associated with T2D susceptibility, we are only beginning to understand the range of T2D allelic variation across global populations at those loci, which genetic processes are functionally modulated, and how they ultimately contribute to T2D pathogenesis. Furthermore, the evolutionary history of established T2D-associated variation has not been fully evaluated. Recent work has suggested that T2D variants harbor extreme levels of population differentiation (1), consistent with the hypothesis that T2D susceptibility alleles may have once been advantageous for regulating metabolic load in early human populations. In order to comprehensively test this hypothesis using all established loci for T2D, and to elucidate global distribution of risk alleles at T2D-associated loci, we computed Weir and Cockerman’s (1984) formulation of FST for 69 LD-independent T2D loci (p<5x10-8) in 1088 unrelated healthy individuals from 14 populations studied by the 1000 Genomes Project (Phase I). We used different combinations of sub-population membership to calculate global FST intra-continental FST (e.g. intra-ASN: CHBxCHSxJPT), and pairwise-continental FST (e.g. AFRxEUR) for each SNP. We computed the empirical distribution of FST from 3.6 M HapMap SNPs for all individuals in each type of FST comparison to examine whether any of the T2D loci fell within the tail of the FST distribution. In contrast to previous observations, our results suggest that as a group, the 69 T2D loci do not have greater FST values than a matched-sampling of SNPs (10,000 random draws) from the rest of the genome (Welch 2-sample t-test, p<0.01). However, upon closer examination of individual T2D loci, we noted one locus showing extreme FST scores across all population comparisons (p<0.01), and exhibiting complex patterns of FST and LD within a Mb of the associated SNP nearby the PTC1 locus. In particular, our investigation revealed at least three LD-independent clusters of high FST, all within the p<0.05 significance cutoff of empirical distribution in the genome (global FST>0.26). We postulate that by triangulating the signals of such LD-independent FST clusters derived from non-coding disease-associated SNPs, we can help identify the causal gene(s) functionally involved in the manifestation of T2D. 1. Chen, et al., PLoS Genetics, 2012.
2016T


The Pima Indians of Arizona have a very high prevalence of type 2 diabetes mellitus (T2DM) and obesity, while Mexican Pima Indians from Maycoba, Sonora, have a much lower prevalence of these disorders. While there are differences in environment between populations, the extent to which they differ at genetic loci for T2DM and obesity is unknown. We genotyped representative population samples of 175 Pimas from Maycoba and 402 from Arizona at 253 putative and established T2DM and obesity variants, along with 96 randomly selected single nucleotide polymorphisms (SNPs). Genetic distance between Mexican and Arizona Pimas, estimated from FST across the random markers, was 0.031 (95% confidence interval 0.019-0.048). Distance across 46 established T2DM variants or across 32 established obesity variants did not differ significantly from that across random markers (FST=0.038, p=0.53, and FST=0.054, p=0.36, respectively). Adjustment for a multiallelic score that sums number of risk alleles across established T2DM variants did not attenuate the increased risk for T2DM in Arizona Pimas, nor did a similar score for obesity attenuate the population difference in body mass index. To identify individual variants that differ between populations, absolute value in allele frequency difference (δf) was calculated and statistical significance was assessed with genomic control, calculated over the random markers, to account for expected effects of population stratification. Five markers showed significant (false discovery rate<0.05) differences in frequency: 4 in HLA-DRB1 (δf=0.56-0.75) and one (rs117619140, δf=0.40, p=3.9×10−7) in Treh. The largest difference was at the HLA-DRB1 variant rs927172 (δf=0.75, p=1.3×10−7). Analysis of data from the Human Genome Diversity Project suggests that this magnitude of allele frequency difference is unusual between human populations of comparable genetic distance: among 55,304,216 allele frequency differences analyzed across 101 pairs of populations with expected effects of population stratification removed, only 25 had a δf larger than the largest difference observed in Mexican and Arizona Pimas. These analyses show that established T2DM and obesity-associated genetic variants explain little of the difference in risk between Mexican and Arizona Pimas, and in general do not show greater divergence than expected based on geographic distance. However, regulatory elements, which have previously been associated with T2DM in Arizona Pimas, have a greater allele frequency differences than expected, perhaps reflecting effects of selection.

2017F

Evolution, adaptation and disease perspectives of Indian populations. S. Nizamuddin1, LVKS. Bhaskar2, P.B. Gai1, BG. Venkatesh2, M. Jyotsna1, K. Thangaraj1. 1) Center for Cellular and Molecular Biology, Hyderabad, Andhra Pradesh, India; 2) Sri Ramachandra University, Porur, Chennai - 600 116, India; 3) Department of Studies in Applied Genetics, Karnataka University, Dharwad-580003, India.

India has been a major corridor for the first modern human, who have migrated out-of-Africa. Since India is inhabited by more than 4,500 anthropologically well-defined populations, who follow strict endogamy marriage practice, they harbor unique genetic architecture compare to populations from rest of the world. India is known not only for its diversity, in terms of social structure, language and culture but also its varied eco-geographical conditions, such as high altitude (Himalayan region: Changkpa population in Kashmir), extreme cold (Himalayan region: Changkpa population in Andhra Pradesh: Mala, Madiga, Vysya). In addition, some populations are living in islands while others are in mainland. Since the Indian populations are inhabited in these geographical regions for tens of thousands of years, we hypothesize these populations might have different selective advantages for certain genetic loci. To fetch signal of natural selection in these population, we have genotyped 134 individual on Affymetrix 6.0 platform and did phasing with Beagle software. Further, we calculated XP-CLH and XP-EHH values, and found novel signal. Interestingly, SFRP1, HIRA, HTR2B, ADD1, FGF1, SIM1, SLC7A10, CEBPA, FASLG and ZFPM2 genes are some examples, which are having importance in survival and genetic fitness for continuing progeny. In addition, we also have found novel genetic factors for many diseases, including cleft palate, breast cancer, etc. Details of these finding would be made available during the presentation.

2018W

Whole genome sequencing reveals past population history and signature of natural selection in a Japanese population. A. Fujimoto1, T. Abe2, K. Boroевич2, K. Nakano1, A. Sasaki1, R. Kidaka1, H. Tanaka2, Y. Nakamura2, S. Miyano2, M. Kubo1, H. Nakagawa1, T. Tsunoda1. 1) Ctr for Integrative Medical Sciences, Riken, Yokohama, Kanagawa, Japan; 2) Human Genome Ctr, The Institute of Medical Science, The University of Tokyo, Japan.

We performed whole genome sequencing of 164 Japanese individuals with 30X coverage, and identified single nucleotide variants, insertions and deletions, copy number variations and rearrangements. We analyzed the pattern of genetic variations and estimated the past population history. Our analysis identified two bottleneck events and a population expansion after the last glacial period. The analysis of the frequency spectrum and the linkage disequilibrium detected signatures of recent positive selection. Significantly higher genetic variation was observed in the HLA and olfactory receptor regions. Our analyses showed the past population history of the Japanese population and identified the signature of adaptation in the local environment of East Asia.

2019T

Integrated natural selection and genome-wide association scan for cholera resistance in Bangladesh. E.K. Karlsson1,2, S. Tabrizi3, D. S. Scalzetti4, T. Struckhoff5,6, Y. Shylakhter, J. D. Desai3, R. Elango2, K. Boroevich1,2, J. B. Harris3, R.C. LaRocque4,5, P. C. Saben1,2,7,1. 1) Broad Institute, Cambridge, MA; 2) Center for Systems Biology, Harvard Univ, Cambridge, MA; 3) International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh; 4) Div. of Infectious Diseases, Massachusetts General Hospital, Boston, MA; 5) Dept of Pediatrics, Harvard Medical School, Boston, MA; 6) Dept of Medicine, Harvard Medical School, Boston, MA; 7) Dept of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA.

As an ancient disease with high fatality, cholera has exerted strong selective pressure on affected human populations, particularly in the Ganges River Delta, where the disease is endemic. This history of selective pressure makes it easier to identify host variants associated with cholera susceptibility and severity. We previously developed a method, the Composite of Multiple Signals (CMS), which combines different tests for natural selection to provide 20-100x better positional resolution than any individual signal. CMS scores each variant, allowing precise identification of the top functional candidates. We have now combined CMS with genome-wide identification to identify selected variants associated with cholera resistance in Bangladesh. We initially performed a genome-wide study of natural selection in a population from the Ganges River Delta and identified 305 candidate selected regions using CMS. The regions were enriched for potassium channel genes involved in cyclic AMP-mediated chloride secretion and for components of the innate immune system involved in NF-κB signaling. Testing the top selected regions in two independent Bengali cohorts found the strongest association with cholera susceptibility in the top region of selection, peaking in the gene SNRNP200. We used these results to develop a model of the human innate immune signaling pathways that respond to V. cholerae infection and have been selected in the Bengali population. In this model, inflammasome activation and the NF-κB signaling pathway play an integrated role in TLR4-mediated sensing of V. cholerae - consistent with our in vitro data. We have now expanded this work by combining a genome-wide association scan for cholera susceptibility with a denser selection scan using full genome sequences for the Bengali population provided by the 1000 genomics project, and have identified new host immune factors likely influencing cholera resistance. Our work shows that using publically available genetic datasets to incorporate tests for natural selection into GWAS analyses can help pinpoint functional variants, an approach applicable to other historically prevalent infectious diseases, such as Lassa fever, tuberculosis, leishmaniasis and malaria, and to complex, common diseases, such as inflammatory bowel disease, for which the associated genes may have been historically selected.
2020F
Targeted analysis of immunogenetic diversity in pre-Columbian Central Andean populations. B. Llamas, G. Valverde-Gamica, A. Cooper, W. Haak. Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, SA 5005, Australia.

Historic accounts estimate that the South American indigenous population size declined by up to 95% between the period of initial contact with Europeans and the beginning of recovery. Most archaeologists and epidemiologists agree that Old World diseases played a major role in this abrupt population collapse. We used advanced ancient DNA techniques to characterize the genetic diversity of loci involved in immune response in pre-Contact Central Andean populations. We have created immortalized ancient DNA libraries using human remains from various periods ranging from the Archaic Period to the Late Horizon. Results from shotgun sequencing showed that nuclear genetic diversity of loci involved in immune response in pre-Contact Central Andean populations.

2020W
Learning Natural Selection from the Site Frequency Spectrum. R. Ronen1, N. Udpa1, E. Halperin2,3,4 V. Bafna1, 1) Bioinformatics and Systems Biology, University of California at San Diego, La Jolla, CA; 2) The Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel; 3) International Computer Science Institute, Berkeley, CA, USA; 4) Department of Molecular Microbiology & Biotechnology, Tel-Aviv University, Tel-Aviv, Israel; 5) Department of Computer Science & Engineering, University of California, San Diego, CA, USA.

Genetic adaptation to external stimuli occurs through the combined action of mutation and selection. A central problem in genetics is to identify loci responsive to specific selective constraints. Many tests have been proposed to identify the genomic signatures of natural selection by quantifying the skew in the site frequency spectrum (SFS) under selection relative to neutrality. We build upon recent work that connects many of these tests under a common framework, by describing how selective sweeps impact the scaled SFS. We show that the specific skew depends on many attributes of the sweep, including the selection coefficient and the time under selection. Using supervised learning on extensive simulated data, we characterize the features of the scaled SFS that best separate different types of selective sweeps from neutrality. We develop a test, SFSselect, that consistently outperforms many existing tests over a wide range of selective sweeps. We applied SFSselect to polymorphism data from a laboratory evolution experiment in which populations of Drosophila melanogaster adapted to hypoxia over hundreds of generations. As a result, we identified loci that support the role of the Notch pathway in hypoxia tolerance. Importantly, several of these Notch-related loci (including the Notch gene region itself) were missed by a previous study. We further applied our test to whole genome sequence data from two human populations. We identified many regions evolving under positive selection, some of which are in agreement with earlier studies, but many of which are novel.

2022T
Reconstructing bacterial phylogeny using whole-genome deep-sequencing data. Y. Lo1, L. Zhang1, B. Foxman2, S. Zoller3,4, 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Epidemiology, University of Michigan, Ann Arbor, MI; 3) Psychiatry, University of Michigan, Ann Arbor, MI.

Studying the genetic variation among bacterial strains from the human body gives insight into the origin and transmission of bacterial infections in human communities, and the emergence and spread of virulence and antibiotic resistance in bacterial populations. While sparse genetic data are sufficient to confidently cluster strains into a `global' phylogeny, a single tree does not fully describe bacterial evolutionary relationships. Ancestries of individual bacterial genes can differ substantially from the global ancestries because of horizontal gene transfer (HGT) and homologous recombination. Comparisons of individual gene trees to the global tree give signals of HGT. Next-generation sequencing methods allow examination of entire bacterial genomes at high resolution and reasonable cost. However, individual gene tree construction often relies on few variant sites on the gene and hence still suffers from lack of data resolution. Here we propose a method to accurately construct and make inferences from gene trees using whole-genome deep-sequencing data of a set of bacterial strains. First we generate the posterior distribution of the global tree using standard Bayesian phylogenetic methods. This global tree posterior distribution serves as prior information for each gene tree. We describe a novel algorithm combining this prior distribution with the likelihood of each gene tree calculated from gene-specific variant calls. Comparing each most probable gene tree to the global tree gives an estimate of the concordance proportion. Bayes factor between each gene tree and the global tree provides a metric for identifying gene phylogenies significantly different from the global phylogeny. We apply the method to a sample of uropathogenic and commensal Escherichia coli strains obtained from human patients with urinary tract infections, each sequenced whole genome at >190x. After constructing global and gene trees for each strain, we find that the global tree posterior distribution serves as a good prior for construction of individual gene trees. The results suggest an ongoing circulation of pathogenic genes, and multiple genetic origins of uropathogenic strains in a small geographic region.

2023F
The adaptive variant EDARV370A is associated with straight hair in East Asians. J. Tan1, Y. Yang1, K. Yang1, P. Sabel1,2,3,4, L. Jin1, S. Wang1,2,3,4, 1) Fudan University, Shanghai, China; 2) Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai, China; 3) The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA.

Hair straightness/curliness is a highly heritable trait among human populations. Previous studies have reported European specific genetic variants influencing hair straightness, but the genetic variants influencing hair straightness in East Asians remain unknown. One promising candidate is a derived coding variant of the ectodysplasin A receptor (EDAR). EDARV370A (370A), associated with several phenotypic changes of epidermal appendages. One of the strongest signals of natural selection in human genomes, 370A has risen to high prevalence in East Asian and Native American populations, while being almost absent in Europeans and Africans. This striking frequency distribution and the pleiotropic nature of 370A led us to pursue if hair straightness, another epidermal appendage-related phenotype, is affected by this variant. By studying 1,718 individuals from four distinctive East Asian populations (Han, Tibetan, Mongolian, and Li), we found a significant association between 370A and the straight hair type in the Han (p = 2.90×10−8), Tibetan (p = 3.07×10−4), and Mongolian (p = 1.03×10−6) populations. Combining all the samples, the association is even stronger (p = 5.18×10−10). The effect of 370A on hair straightness is additive, with an odds ratio of 2.05. The results indicate very different biological mechanisms underlying straight hair in Europe and Asia, and also put hair in a more comprehensive picture of the phenotypic consequences of 370A, providing important clues into the potential adaptive forces shaping the evolution of this extraordinary genetic variant.
2024W
Macrophages from African and European populations respond differently to bacterial infection. Y. Nédélec1,2, A. Pagé Sabourn1,2, V. Yotova1, J.C. Grenier1, N. Cotta1, L.B. Barreiro1,2, 1) Research Center, CHU Sainte-Justine, Montréal, Canada; 2) University of Montreal, Montréal, Canada.

Infectious diseases have always been a major health problem throughout the world, imposing strong selective pressure on the human genome. Geographically distinct human populations are postulated to have differing histories of pathogen exposure. Indeed, previous studies demonstrate that people of African and European ancestry differ in their susceptibility to certain infectious diseases like tuberculosis, malaria and sepsis. Differences in infection progression between African and European populations suggest inter-population variation in the immune response, possibly caused by the adaptation of Africans and Europeans to the pathogens of their environment. For the first time, we characterize the immune response of people of African and European ancestry to bacterial infections. We infected monocyte-derived macrophages from 24 African Americans and 24 European Americans with the intracellular pathogens Listeria monocytogenes and Salmonella typhimurium for 4 hours and measured whole genome gene expression of infected and non-infected cells by RNA-sequencing. We assessed macrophage control of bacterial infection at 1 hour and 24 hours by culturing infected cell lysate and counting colony-forming units to approximate the rate of bacterial survival. We found that macrophages derived from people of African ancestry presented fewer intracellular bacteria after 24 hours than people from European ancestry, suggesting that the African Americans better control intracellular bacterial infections. Concordantly with this observation we identified inter-population differences in immune gene pathway expression that might explain this pattern of increased infection control in African American macrophages. Interestingly, multiple genes up-regulated by bacterial infection in people of European ancestry were found to be already highly expressed in the non-infected cells of people from African ancestry. We show that several of these genes appear to have been subject to recent selection, which may explain between population disparities in their expression. Our study identifies multiple candidate genes that may affect the course of L. monocytogenes and S. typhimurium infection in humans. Importantly, our findings suggest that the clinical differences in infectious disease progression observed in populations of African and European ancestry may be outcome of natural selection.

2025T
Co-evolution of HLA class I with killer cell immunoglobulin-like receptors in a sub-Saharan African population. P.J. Norman1, J.A. Hollenbach2, N. Nemati-Gorgani1, L.A. Guethlein2, H.O. Hilton1, M.J. Pandolfi2, K. Koram2, E.M. Riley3, L. Abi-Rached4, P. Parham1, 1) Department of Structural Biology, Stanford University School of Medicine, Stanford, CA; 2) Center for Genetics, Children’s Hospital Oakland Research Institute, Oakland, CA; 3) Department of Pathology, Stanford University School of Medicine, Stanford; 4) Noguchi Memorial Institute for Medical Research, University of Ghana, Legon, Ghana; 5) Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK; 6) Centre National de la Recherche Scientifique, Laboratoire d’Analyse, Topologie, Probabilités, Équipe ATIP, Aix-Marseille Université, Marseille, France.

Human natural killer (NK) cell and cytotoxic T cell responses to pathogens, and the role of NK cells in reproduction, are controlled by diverse receptors that interact with highly polymorphic HLA class I molecules. Although T-cell receptor (TCR) selection is generated somatically with little genetic determination, killer cell immunoglobulin-like receptors (KIR) are highly polymorphic and polygenic and specific combinations of HLA class I and KIR alleles are differentially associated with numerous disease susceptibilities and therapeutic outcomes. Extensive KIR gene-content variation is observed between human populations. However, high-resolution analysis of KIR allele and haplotype diversity has focused on Asian and Amerindian populations having limited genetic diversity and the more heterogeneous sub-Saharan African populations have remained under-represented. We studied KIR and HLA variation in 235 individuals, including 104 mother-child pairs from the Ga-Adangbe ethnic group of Ghana in West Africa. This population has a rich diversity of 175 HLA haplotypes and 81 HLA-A, -B and -C variants that form 190 HLA class I haplotypes. Balancing selection has maintained this exceptional polymorphism and diversified interaction of HLA class I with KIR, with evidence for their ongoing co-evolution in the Ga-Adangbe. Highlighting the critical role in countering pathogen diversity, every individual studied (N=188) has a unique compound genotype of KIR and HLA class I. Nevertheless, comparison of the observed frequencies with those simulated under a model of balancing selection revealed the impact of positive selection that has reduced the diversity of functionally interacting sites. Whereas the centromeric region of the Ga-Adangbe KIR locus is exceptionally diverse, the telomeric region lacks diversity due to the low frequencies of Tel B KIR genes and alleles and correspondingly high frequencies of Tel A KIR genes and alleles. As a consequence of the high frequency of HLA-B*53, -B*35 and related allotypes in the Ga-Adangbe, diversity is reduced at residues in the α1 domain of HLA-B that interact with the variable region of T-cell receptor β chains. These features of KIR and HLA-B are consistent with selection by a pathogen endemic to West Africa. The previously reported association of HLA-B*53 with severe malaria and the high prevalence of malaria in the Ga-Adangbe population strongly suggest Plasmodium falciparum as a candidate.
Host Genetics and Human Adaptation to Lassa Hemorrhagic Fever in West Africa. S. Tabrnà1,2,3,10, R. Taryal1,2,10, I. Shiyakhzner1,2, S.F. Schaffner1,2, S.K. Gire1,2, M.S. Stromlau1,2, E.K. Karlsson1,4, K.G. Andersson1,2, E. Phelan1, L.M. Moses3, W. Omoninya1, I. Odia2, P.E. Ehiane3, O. Folaru5,6, A. Tarebj1,2, L.M. Branco3, J.S. Schieffelin4, D. Levy5,7, S. Gunther5, D.S. Grant1, G.O. Akpede2, D.A. Asogun3, P.O. Okokhere6, R.F. Garry1,4, C.T. Happel6,8,11, P.C. Sabel1,2,2,11, Viral Hemorrhagic Fever Consortium. 1) FAS Center for Systems Biology, Department of OEB, Harvard University, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) Harvard Medical School, Boston, MA; 4) Department of Microbiology and Immunology, Tulane University, New Orleans, LA; 5) Institute of Lassa Fever Research and Control, Irrua Specialist Teaching Hospital, Edo State, Nigeria; 6) Department of Biological Sciences, College of Natural Sciences, Redeemer’s University, Redemption City, Ogun State, Nigeria; 7) Lassa Fever Laboratory, Kenema Government Hospital, Kenema, Sierra Leone; 8) Lassa Fever Laboratory, Kenema Government Hospital, Kenema, Sierra Leone Department of Virology, Bernhard-Nocht-Institute for Tropical Medicine, Hamburg, Germany; 9) Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA; 10) These authors contributed equally to this work; 11) These authors jointly supervised this work.

Lassa virus has the unique status of being both one of the world’s deadliest pathogens, designated bio-safety level 4 (BL-4), and a public health crisis, endemic in large parts of West Africa. It is estimated to have originated over 1000 years ago in Nigeria, and a genome-wide survey of human variation identified the gene LARGE, biologically linked to Lassa virus infection, as among the strongest signals of natural selection in the Yoruba population of Nigeria. We pursued the hypothesis that Lassa virus is an ancient selective force driving the strong positive genetic resistance, by conducting a genome-wide association study (GWAS) of Lassa fever cases and controls in Nigeria and Sierra Leone. We report a signal of association at LARGE, and show that it overlaps with the previously reported signal of selection. This suggests strong positive selection pressures and result in recent positive selection events. Beyond Nigeria, and helps to explain higher fatality rates and greater disease severity in Sierra Leone. Beyond LARGE, our other top GWAS signals fall at genes involved in the immune system, including PTPRE, which modulates expression of cytokines involved in host response to viral infection, and IFRD1, involved in the activation of neutrophils in innate immune response. We also examine signals of adaptation in the Esan and Mende populations and discover numerous novel candidate loci under positive selection, providing new insights into human diversity in West Africa.

The study of positive selection in human genetic variation provides unique insight into human diversity in West Africa. A GWAS of Lassa fever cases and controls in Nigeria and Sierra Leone identified a signal at the LARGE gene, which is involved in the innate immune response. This signal overlaps with previously reported signals of selection and suggests strong positive selection pressures in recent evolution of resistance to Lassa fever. The study also identifies additional signals in genes involved in the immune system, providing new insights into human diversity in West Africa.
Identification of regions under positive selection in the Gullah African American population of South Carolina. P.S. Ramos1, S.P. Sajuthi1, J. Divens2, Y. Huang3, U. Nayak4, W.M. Chen5, K.J. Hunt1, D.L. Kamen7, G.S. Gilkeson1, J.K. Frendanes1, J.I. Spruill2, C.D. Lange1, W.T. Garvey5, M.M. Sale6. 1) Medical University of South Carolina, Charleston, SC; 2) Wake Forest School of Medicine, Winston-Salem, NC; 3) University of Virginia, Charlottesville, VA; 4) University of Alabama, Birmingham, AL.

The Gullahs form a unique population of African ancestry in the U.S. In addition to their relative genetic and environmental homogeneity and low European admixture (less than 11% on average) a shorter genetic distance between the Gullah and Sierra Leoneans has also been noted, suggesting that population genetic signals, such as regions under recent selection, may be more easily detected in the Gullah than in other African American (AA) populations. Given the increased prevalence of some complex diseases in AA and the increasing evidence of selection at loci associated with human diseases, identification of alleles under selection may provide insight into disease susceptibility. Since population-specific selection may cause allele frequency differences, the goal of this study was to identify regions with minor allele frequency (MAF) differences between Gullah and Sierra Leoneans. We had available 277 Gullah and 400 Sierra Leonian samples genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0. After stringent QC, 679,513 SNPs with MAF>5% were used to compute the significance of the MAF differences between the two populations. The European component was adjusted for via inclusion of the HapMap CEU principal component in the logistic regression model. In order to exclude spurious MAF differences, only regions where at least two SNPs in LD showed suggestive (P<10-05) MAF differences were considered. The region showing the most significant MAF difference between the Gullah and Sierra Leoneans was at 20q13.13, where multiple SNPs in LD showed higher MAFs in the Gullah and P-values between E-03 and E-07. As shown in the HaploTagger and HGDP Selection Browsers, this region shows evidence for selection in the Mandenka, who are geographically the closest to Sierra Leone. In summary, we have identified a ~500 kb region of the telomeric segment of the KIR region that is independent of demographic history.

Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. B. Su1, K. Xiang1,6, NA. Ou Zhulabu2, Y. Peng1, Z. Yang1,6, X. Zhang1,6, C. Cui1, H. Zhang1, M. Li1,6, Y. Zhang1, NA. Bianba2, NA. Goggolanzi3, NA. Basang4, NA. Ciwang-sangbu5, T. Wu1, H. Chen1, H. Shi1, X. Qi1. 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Inst Zoology, Chinese Academy Sci, Kunming, Yunnan, China; 2) High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa, China; 3) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining, Qinghai, China; 4) People’s Hospital of Dangxiong County, Danxiong, China; 5) Department of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, USA; 6) University of Chinese Academy of Sciences, Beijing, China.

Tibetans are well adapted to high-altitude hypoxic conditions, and in recent genome-wide scans, many candidate genes have been reported involved in the physiological response to hypoxic conditions. However, the limited sequence variations analyzed in previous studies would not be sufficient to identify causal mutations. Here we conducted re-sequencing of the entire genomic region (59.4 kb) of the hypoxic gene EGLN1 (one of the top candidates from the genome-wide scans) in Tibetans, and identified 185 sequence variations including 13 novel variations (12 substitutions and one in-del). There is a non-synonymous mutation (rs186996510, D4E) showing surprisingly deep divergence between Tibetans and lowlander populations (FST = 0.709 between Tibetans and Han Chinese). It is highly prevalent in Tibetans (70.9% on average), but extremely rare in Han Chinese, Japanese, Europeans and Africans (0.56%-2.27%), suggesting that it might be the causal mutation of EGLN1 contributing to high-altitude hypoxic adaptation. Neutrality test confirmed the signal of Darwinian positive selection on EGLN1 in Tibetans. Haplotype network analysis revealed a Tibetan-specific haplotype, which is absent in other world populations. The estimated selective intensity (0.029 for the C allele of rs186996510) puts EGLN1 among the known genes undergone the strongest selection in human populations, and the onset of selection was estimated to have started at the early Neolithic (~8,400 years ago). Finally, we detected a significant association between rs186996510 and hemoglobin levels in Tibetans, suggesting that EGLN1 contributes to the adaptively low hemoglobin level of Tibetans compared to acclimatized lowlanders at high altitude.

2029F Fluctuating And Geographically Specific Selection Characterize Rapid Evolution Of The Human Killer Immunoglobulin-Like Receptor (KIR) Locus. J.A. Hollenbach1, P.J. Norman2, J.J. Rutter3, E.A. Trachtenberg1, P. Parham1, K.D. Taylor4. 1) Center for Genetics, Children’s Hospital Oakland Research Institute, Oakland, CA; 2) Departments of Structural Biology and Microbiology & Immunology, Stanford University School of Medicine, Stanford, CA; 3) Medical Genetics Institute, Cedars Sinai Medical Center, Los Angeles, CA.

Killer cell immunoglobulin-like receptors (KIR) are expressed on human natural killer (NK) cells and a small percentage of cytotoxic T-cells, and regulate immune responses during infection. Diversity of KIR is known to impact NK and T cell function and has been associated with infectious and autoimmune diseases, as well as cancer and transplant outcome. The KIR gene complex of human chromosome 19q13.4 displays extensive allelic and gene-content variability, with between 7-14 highly polymorphic genes. To examine KIR evolution in the context of the entire human genome, gene-content diversity and 125 SNPs in the KIR and flanking regions were compared to ~650,000 genome-wide SNPs in 852 individuals from 52 populations of the human genome diversity panel (HGDP). KIR allelic diversity was further examined in a subset of 60 individuals. In both African and Oceanic populations, SNP diversity and linkage disequilibrium (LD) patterns in the telomeric segment of the KIR region show strong evidence for purifying selection in response to local pressures. Prior to emergence of modern humans, KIR3DL1/S1 diverged into three lineages that encode high and low expressing inhibitory or activating receptors, respectively. In Africa, ongoing selection specifically favors the highly expressing inhibitory receptors. In contrast, in Oceanic populations there is evidence of strong and recent significant MAF differences between the receptor and Sierra Leonian alleles of the KIR region, high diversity worldwide is consistent with balancing selection, particularly in Oceania. Highlighting the considerable fluctuation of selection pressures, East Asians exhibit ongoing purifying selection on the activating KIR that began before divergence of Amerindians; populations in which centromeric KIR diversity subsequently rebounded under balancing selection. In all populations exons 7-9 of KIR3DL3 and KIR3DL2 mark the centromeric and telomeric boundaries of LD within the KIR region.
2032F
Inference of Natural Selection and Demographic History for African Pygmy Hunter-Gatherers. P.H. Hsieh1, K.R. Veeramah1, J. Lachance4, S.A. Tishkoff1, J.D. Wall3, M.F. Hammer1,2, R.N. Gutenkunst1,2, 1) Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ; 2) Arizona Research Laboratories, Division of Biotechnology, University of Arizona, AZ; 3) Department of Molecular and Cellular Biology, University of Arizona, AZ; 4) Department of Biology and Genetics, University of Pennsylvania, Philadelphia, PA; 5) Institute for Human Genetics, University of California, San Francisco, CA.

African Pygmies are hunter-gatherers primarily inhabiting the Central African rainforests, where they are exposed to high temperatures, high humidity, and a pathogen and parasite-enriched woody habitat. These factors undoubtedly influenced their evolutionary history as they adapted to this environment. Many Pygmy populations have historically been in socio-economic contact with neighboring Niger-Kordofanian speaking farmer populations, particularly since the agriculture expansion in sub-Saharan Africa that began five thousand years ago (kyr). To look for the true signatures of adaptation to the rainforest habitat of pygmies we must control for this complex demographic history. We sequenced and combined 40x whole genome sequence data from 3 Baka pygmies from Cameroon, 4 Baka pygmies from the Central African Republic, and 9 Niger-Kordofanian speaking Yoruba farmers from Nigeria. We used iaii, a model-based demographic inference tool, to infer the history of these populations. Our best-fit model suggests that the ancestors of the farmer and pygmy populations diverged 150 kyr and remained isolated from each other until 40 kyr. This divergence is more ancient than estimated by previous studies that included fewer loci, but is consistent with a PSMC analysis, a separate inference tool that uses different aspects of the genomic data than iaii. Interestingly, our analysis shows that models with bi-directional asymmetric gene flow between farmers and pygmies are statistically better supported than previously suggested models with a single wave of uni-directional migration from farmers to pygmies. To identify possible targets of positive selection, we conducted a genomic scan using complementary methods, including the frequency-spectrum based G2D test, the population differentiation based XP-CLR test, and the haplotype based iHS test. We performed 10,000 simulations based on the admixed population. As we expected, we identified a genomic region of lipid binding are enriched in highly differentiated non-synonymous mutations, suggesting that this function may have acted differently on the Pygmies and farmers after their divergence from their common ancestor.

2034T
Variation in Bone Mineral Density (BMD) in Children of Different Ethnic Backgrounds is Explained By Genetic Profiling: The Generation R Study. M. Medina-Gomez1,2,3, D.H.M. Hoppa2,4, K. Estrada1,2,6,5, E. Oei1,2, A. Hofman1,2,3, M. Kayser1, A.G. Uitterlinden1,2,3, V.W. Jaddoe2,4,5, O. Lao6, F. Rivadeneira1,2,3, 1) Internal Medicine, Erasmus MC University, Rotterdam, Netherlands; 2) Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 3) The Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands; 4) Pediatrics, Erasmus Medical Center, Rotterdam, The Netherlands; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 7) Harvard Medical School, Boston, MA, USA; 8) Forensic Molecular Biology, Erasmus Medical Center, Rotterdam, The Netherlands.

Aim: To study genetically-determined differences in BMD in children across ethnicities, we constructed a genetic score (GS) of BMD SNPs in children from multiple ethnic background and tested for association with total-body (TB-) BMD. Methods: We included 4,009 children from The Generation R study, a single-center prospective multiethnic birth cohort in Rotterdam, The Netherlands, with GWAS data and TB-BMD measured by DXA at a mean age of 6.2 years. The admixed population includes a wide spectrum of ethnic origins including blends between continental groups. Clustering based on genetic data (using ADMIXTURE) was employed assigning children to one of three transcontinental ancestral groups: predominantly Caucasian (n=3,513), Asian (n=159) and African (n=337). Differences in TB-BMD were assessed by least-squares means using the Caucasian population as reference and adjusting for age, sex, height, fat and lean mass. The GS consisted of the sum of BMD-increasing alleles (0,1,2) in genetic models identified in Affymetrix Axiom array and analyzed across quintiles. Results: Children of African descent had higher TB-BMD (0.585 g/cm2) than those of Asian (0.559 g/cm2; P=2E-6) and Caucasian descent (0.552 g/cm2; P=1E-6). The GS explained 5.2% of the variance in BMD and 2.6% after principal component correction. As compared to children in the GS middle quintile (54% of the population; n=2,179), individuals in the GS highest quintile (2.6% of the population; n=106) had 0.69 SDs higher BMD (P<1E-6), while those in the GS lowest quintile (1.9% of the population; n=177) had 0.53 SDs lower BMD. The number of children of African descent were overrepresented (P=3E-104) in the two highest quintiles of the GS with 65.6%, as compared to 18.8% in children of Caucasian and 15.7% in those of Asian descent. Positive differences of >10% in BMD-increasing allele frequencies were seen between children of African and Caucasian descent for 21 variants and differences of > 5% for 34 variants of the score. Data from the Human Genome Diversity panel also indicated higher frequency of BMD-increasing alleles in African populations, with South-west Bantuans displaying the highest frequencies. Conclusions: Genetic ancestry of African descent is correlated with a higher frequency of BMD-increasing alleles. Ethnic differences in bone accrual are already evident at childhood and are partially explained by allele frequency differences in variants from known BMD loci.

2033W
Forward simulations of recurrent selection and demographics with rescaled parameters. L.H. Unichio1, R.D. Hernandez1, 1) Bioengineering Graduate Group, UC Berkeley & UCSF, San Francisco, CA; 2) Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA.

It is well known that positive selection impacts patterns of diversity in linked regions of a genome. When a population experiences frequent positive selection, the resulting patterns of genetic variation in linked regions can be quite skewed, but theoretical treatments of this process (often referred to as ‘recurrent hitchhiking’) are mostly limited to simple models. Forward simulation can provide insights into more sophisticated models that include a number of genetic and demographic factors, including the computational burden of forward simulation is often prohibitive. A remedy to this computational challenge is to rescale the relevant parameters (e.g., population size) in a way that conserves the underlying dynamics. However, ad hoc approaches to parameter scaling in recurrent hitchhiking may not always provide sufficiently accurate dynamics, potentially skewing patterns of diversity in simulated DNA sequences. Here, we perform a detailed theoretical analysis of the recurrent hitchhiking model that relaxes some simplifying assumptions, and present a simple method for parameter rescaling under the model. We thoroughly test the robustness of rescaling across the parameter space, and describe the conditions under which rescaling provides accurate results. We find that our approach enables us to perform large-scale simulations in a fraction of the computational time. We apply this rescaling method to simulations of interference among selected sites as well as demographic models of growth and contraction.
2035F
Non-Random Geospatial Distributions of HLA-Haplotypes in the United States. N.P. Leahy, A. Chatterjee, M. Abrecht, M. Maiers. 1) Biometrics Research, Be The Match, Minneapolis, MN; 2) School of Statistics, University of Minnesota, Minneapolis, MN.

A key predictor of hematopoietic cell transplantation success is using a donor with a fully HLA-matched donor for the patient. Donor registries such as the Be The Match® registry require knowledge of human population structure to target recruitment to subpopulations with rare variants. We used data from the Be The Match® registry to test for spatial heterogeneity within self-identified race and ethnic groups (SIRE). In total there were four broad SIRE groups (African-American, Asian/Pacific Islander, Caucasian, Hispanic) and 18 subcategories of the broad SIRE groups. The high genetic variability at HLA and low haplotype frequencies presented challenges for traditional analysis methods. Therefore, we opted for a permutation methodology to establish heterogeneity. For each of the SIRE groups, we generated a matrix, $H$, of the occurrences of the top 100 haplotypes in 50 US states. District of Columbia, and Puerto Rico. Our metric was the maximum eigenvalue for the variance-covariance matrix, $\lambda_{\text{max}}$. To test for spatial heterogeneity in our data, we performed 10,000 permutations of the data to generate a probability distribution of $\lambda_{\text{max}}$. Significance was determined by the position of the metric to the distribution. This was repeated for the top 100 haplotypes for 1000 permutations. The permutation analysis found for the 100 most common haplotypes, Asian/Pacific Islander ($p < 0.0001$) and Hispanic ($p = 0.0036$) had significant special heterogeneity and Caucasian was marginally insignificant. When the analysis was repeated with the 1000 most common haplotypes, Caucasian was not significant ($p < 0.0001$) and Hispanic was no longer significant ($p = 0.412$). Outcomes of the subcategories mostly reflected those of the broad SIRE groups. African-American was not significant. These results were consistent with historical founder events. That not all SIREs can be treated as homogenous subpopulations has implications for recruitment strategies of donor registries. To better serve patients, it is necessary to resolve the patterns within spatially heterogeneous SIRE groups to maximize diversity within registries. That increasing the number of haplotypes shifted some SIRE groups from insignificant to non-significant is problematic. The expectation was adding additional genetic information should have provided better resolution of population structure, requiring a more nuanced hypothesis.

2036W
Proportion of African ancestry in Helicobacter pylori is associated with increased severity of gastric lesions in human hosts with high Amerindian ancestry. N. Kodaman, T. Boszo, T. Schneiter, S. Sobota, M.B. Piazuelo, C.L. Shaffer, J. Romero-Gallo, T. de Sablet, L.E. Bravo, K. Wilson, T. Cover, S.M. Williams, P. Correa. 1) CHGR, Vanderbilt University, Nashville, TN; 2) Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Pathology, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Pathology, Universidad del Valle School of Medicine, Cali, Colombia; 5) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN; 6) Division of Infectious Diseases, Vanderbilt University Medical Center, Nashville, TN; 7) Veterans Affairs Tennessee Valley Healthcare System, Nashville, TN; 8) Universidad de Nariño, Pasto, Colombia; 9) Department of Genetics, Dartmouth University, Hanover, NH. Helicobacter pylori is the principal cause of gastric cancer, the second leading cause of cancer mortality worldwide. In certain Colombian populations, over 90% of individuals are infected with $H.\, pylori$, but infection rate does not generally predict cancer prevalence. In particular, residents of the Andean mountain region are 25 times more likely to develop gastric cancer than their coastal counterparts, despite similar rates of infection. We determined the ancestry of $H.\, pylori$ isolates from the gastric biopsies of 275 Colombian subjects from both the mountain and coastal regions. Most isolates contained genomic regions from four ancestral $H.\, pylori$ populations: Africa1 (AA1), Europe1 (AE1), Europe2 (AE2), and East Asia (AEA), but these proportions varied with geography. The AA1 cluster was more common in coastal samples (mean=47.9%), and AE2 in mountain samples (mean=50.7%). The human ancestry of the biopsied individuals also varied with geography. With mean proportion of American ancestry, 57.9% African, 22.6% Amerindian, and 19.5% European ancestry in the coastal region, and 67% Amerindian and 30.4% European ancestry in the mountain region. All pairwise correlations between $H.\, pylori$ ancestry and human host ancestry were significant. While African and $H.\, pylori$ ancestry correlated negatively with Amerindian ancestry, and $H.\, pylori$ ancestry per se was not responsible for the increased severity of the lesions. Importantly, region of origin (coast or mountain) was also not significant when added as a covariate to the mutivariate model ($p=0.89$). Our findings indicate that AA1 ancestry is relatively benign in human ancestry, but potentially deleterious in Amerindians, making the interaction between human and $H.\, pylori$ ancestry a potential confounder of univariate analyses on histopathology. Co-evolution has likely modulated disease risk, and the disruption of this relationship may account for the risk discrepancy in Colombian populations.

2037T
Principal component analysis reveals the 1000 Genomes Project does not sufficiently cover the human genetic diversity in Asia. D. Lu, S. Xu. Partner institute for Computational Biology, Shanghai, China.

The 1000 Genomes Project (1KG) aims to provide a comprehensive resource on human genetic variations. With an effort of sequencing 2,500 individuals, 1KG is expected to cover the majority of the human genetic diversities worldwide. In this study, using analysis of population structure based on genome-wide single nucleotide polymorphisms (SNPs) data, we examined and evaluated the coverage of genetic diversity of 1KG samples with the available genome-wide SNP data of 3,831 individuals representing 140 population samples worldwide. We developed a method to quantitatively measure and evaluate the genetic diversity revealed by population structure analysis. Our results showed that the 1KG does not have sufficient coverage of the human genetic diversity in Asia, especially in Southeast Asia. We suggested a good coverage of Southeast Asian populations be considered in 1KG or a regional effort be initialized to provide a more comprehensive characterization of the human genetic diversity in Asia, which is important for both evolutionary and medical studies in the future.
2038F

The Iranian Genomes Project. R. Daneshjou1, M. Ronagh2, C.D. Bustamante1, P.C. Sabel3, R.B. Altman1. 1) Genetics, Stanford University, Stanford, CA; 2) Illumina Inc., San Diego, CA; 3) Department of Human Evolutionary Biology, Harvard University, Cambridge, MA.

While large-scale full-genome studies, such as the 1000 Genomes Project, have examined genetic variation for many of the world’s populations, Middle Eastern populations have remained little studied. This group’s unique history and geographic location, at the presumed bottleneck of the out of Africa migration, make them an important population to investigate. Here, we present the Iranian Genome Project, an effort to deeply sequence (at 30X coverage) over 50 individuals of Iranian descent. Our cohort includes conservatively and understudied sub-populations from Iran, such as Zoroastrians, Bakhhtiari, and Jewish populations. We explore the relationship between the Iranian population to other 1000 Genomes populations and the relationships between the Iranian sub-populations using EIGENSOFT software and the HAPMIX algorithm to compute FST and Principle Components Analysis (PCA). We further identify variants that are unique to the Iranian population and predict the functional impact of these variants. Because ancestry is an important consideration in clinical genetics, we identify the frequency of known or predicted damaging mutations in genes identified by the recent American College of Medical Genetics and Genomics (ACMG) guidelines on incidental findings in genome sequencing. Additionally, we calculate frequencies for important known clinical pharmacogenetic variants and identify new variants in known pharmacogenetic genes, which may play an important role in this population. Given the importance of diverse reference sequences for both population and clinical genetics, the Iranian Genome Project serves to fill in the gap by providing deeply sequenced genomes from a key but unexplored Middle Eastern population.

2039W

Morphometric and ancient DNA study of human skeletal remanants in Indian Subcontinent. N. Rai1, M. Mirazon Lahr2, L. Singh1, K. Thangaraj. 1) Evolutionary and Medical Genetics, Centre for Cellular and Molecular Biology, Hyderabad, India; 2) Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, U.K.

Recovery and sequencing of mtDNA from ancient human remains is a daunting task but provides valuable information about human migrations and evolution. Our present study is the first to recover, amplify and sequence (HVR and coding regions of mtDNA) inadequately preserved and highly degraded (1.5 Ky to ≤1.0 Ky ago) hominids mitochondrial DNA of three most intriguing and indigenous ancient population of South and South-East Asia (Myanmar=20 Buried individuals, Nicobar Islands=15 and Andaman Island=6). Following all parameters and to avoid the chance of contamination we independently extracted and sequenced the DNA in two different labs and measured the cranial variability in all hominid skulls using 128 cranial landmarks, compiled 3D morphometrics, genetic data of ancient DNA samples and analyzed the admixture and genetic affinities of above three populations. Results showed the predominant frequency of F1a1 and complete absence of 9bp deletion in ancient Nicobares. Unlike in previous reports on modern Nicobarese, the high frequency of F1a1 haplogroup in ancient Nicobarese show the probable migration of Nicobarese from South East Asia and the complete absence of 9bp deletion suggests the different events of settlement. This study failed to detect genetic affinities of Burmese with Nicobarese even though their phenotype and language appears to be same. We first time report any kind of population study on Burmese populations and with the genetic affinity of Burmese with East Asian, East Indian (Including Gadhwal region of Himlaya) and Bangladeshi populations, we found significant admixture with West Eurasians. Our study strongly supports the West Eurasian and East Asian route of migration and settlement of early Burmese population. The three populations in the present study are quite different in their genetic structure but 3D morphometric study using huge number of landmarks explains a close homology among these populations and this can be explained by the role of climatic signature on these populations.

2040T

No significant differences in the accumulation of deleterious mutations across diverse human populations. R. Do1,2, D. Balick1,2, J. Aadhubey1, S. Synayev1,2, D. Reich1,2. 1) Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142; 2) Center for Human Genetics Research, Massachusetts General Hospital, Boston, Massachusetts, 02114; 3) Division of Genetics, Brigham & Women’s Hospital, Harvard Medical School, Boston, Massachusetts, 02115; 4) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115.

Differences in demographic history across populations are expected to cause differences in the accumulation of deleterious mutations because natural selection works less efficiently when population sizes are small. Surprisingly, however, the relative burden of deleterious mutations has never been directly measured across human populations on a per-haploid genome basis, despite the fact that this is what matters biologically in the absence of dominance and epistasis. Here we empirically measure the relative accumulation of deleterious mutations in 13 diverse populations (Yoruba, Mande-nka, San, Mbuti, Dinka, Australian, French, Sardinian, Han, Dai, Mbe, Karitiana and Papuan) along with one archaic population (Denisova). All the present-day populations have statistically indistinguishable accumulations of coding mutations. We highlight two examples. First, we find no evidence for a lower mutational load in West Africans than in Europeans despite the approximately 30% higher genetic diversity in West Africans; the accumulation of non-synonymous mutations in West Africans is 1.01±0.02 times that in Europeans, and for “probably damaging” mutations, the ratio is 1.03±0.04. Second, we find no evidence for a lower mutational load in populations that have experienced agriculture-related expansions over the last 10,000 years and those that have not: the ratio in Chinese to Karitiana hunter gatherers from Brazil is 0.99±0.07. We determined that these null results are not an artifact of insensitivity of our method to differences in demographic history. As a positive control, we also analyzed archaic Denisovans who are known to have had a small population size for hundreds of thousands of years since separation from modern humans. We show that the Denisovan lineage has accumulated “probably damaging” mutations 1.33±0.06 times more rapidly than modern humans since they split. These analyses are important because of the new constraints they place on the distribution of selection coefficients among humans. Given the histories of West Africans and Europeans, combined with the fact that we do not detect a lower accumulation of deleterious mutations in West Africans than Europeans, we can conclude that only a small proportion of non-synonymous mutations since separation from modern humans. These findings are consistent with a large number of previous studies on human populations that have experienced rapid population expansions over the last 10,000 years. We further identify variants that are unique to the Iranian population and predict the functional impact of these variants. Because ancestry is an important consideration in clinical genetics, we identify the frequency of known or predicted damaging mutations in genes identified by the recent American College of Medical Genetics and Genomics (ACMG) guidelines on incidental findings in genome sequencing. Additionally, we calculate frequencies for important known clinical pharmacogenetic variants and identify new variants in known pharmacogenetic genes, which may play an important role in this population. Given the importance of diverse reference sequences for both population and clinical genetics, the Iranian Genome Project serves to fill in the gap by providing deeply sequenced genomes from a key but unexplored Middle Eastern population.

2041F

Y chromosomes of ancient Hunnu people and its implication on the phylogeny of East Asian linguistic families. LL. Kang1,2, TB. Jin1, F. Wu1,2, X. Ao2, SQ. Wen2, CC. Wang2, YZ. Huang2, XL. Li1,2, H. Li1,2. 1) Key Laboratory of High Altitude Environment and Gene Related to Disease of Tibet Ministry of Education, School of Medicine, Tibet University for Nationalities, Xianyang, Shaanxi, China; 2) MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China.

The Hunnu (Xiongnu) people, also called Huns in Europe, were the largest ethnic group to the north of Han Chinese until the 5th century. The linguistics, affilations of the Hunnu are controversial among Yeniseian, Altaic, Uralic, and Indo-European. Recent DNA analyses on the remains of the Hunnu people showed some clues to this problem. Y chromosome haplogroups of Hunnu remains included Q-M242, N-Tat, C-M130, and R1a1. Recently, we analyzed three samples of Hunnu from Barköl, Xinjiang, China, and determined Q-M3 haplogroup. Therefore, most Y chromosomes of the Hunnu samples examined by multiple studies are belonging to the Q haplogroup, and most Y chromosomes in the Hunnu people should be in the Yeniseian family. The Y chromosome diversity is well associated with linguistic families in East Asia. According to the similarity in the Y chromosome profiles, there are four pairs of congeneric families, i.e., Austroasiatic and Tai-Kadai, Mon-Khmer and Hmong-Mien, Sino-Tibetan and Uralic, Yeniseian and Palaisiberian. Between 0,000-2,000 years before present, Tai-Kadai, Hmong-Mien, Sino-Tibetan, and Yeniseian languages formed into distinct languages, becoming quite different from the rest four. Since Hunnu was in the Yeniseian family, all these four languages were distributed in the north of China during the transformation. There must be some social or biological factors induced the transformations at that time, which is worth doing more linguistic and genetic researches.
Admixture in the Pre-Columbian Caribbean, J.C. Martinez-Cruzado1, E.P. Tascón-Peraranda1, F. Curbelo-Canabal1, T. Porral-Doria1, C. Eng2, E.G. Burchard2 1) Department of Biology, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico; 2) Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco, San Francisco, California.

The biological origin of the Caribbean aborigines that greeted Columbus is one of the most controversial issues regarding the population history of this region. Genome studies suggest an Equatorially-Tucanoan origin, consistent with the Arawakan language spoken by most natives of the region. However, the archaeological evidence suggests an early arrival from Mesoa-america, and their admixture with the more recent Arawak-speaking groups stemming from the Amazon remains a possibility. The lineages comprehending most Puerto Rican samples belong to haplogroups B1 and C1, which in turn encompass 44% of all Native American mtDNAs, have been related to South America or any other continental region. To augment the scarce data from Mesoamerican countries other than Mexico, we present the mtDNA sequence data of six Honduran samples belonging to distinct control region lineages in addition to 3 from the Dominican Republic and 3 from Puerto Rico. Interestingly, maximum likelihood phylogenetic recon- struction including 40 published haplogroup A2 sequence haplotypes from Mesoamerica, Central America and South America clusters 8 out of 10 Mesoamerican and Andean haplotypes in a deep rooted group, separate from, and excluding all Costa Rican, Panamanian and Brazilian haplotypes, suggesting a relatively recent origin for Chibchan-Paezan and Amazonian groups. Furthermore, 4 of the 5 Greater Antillean A2 haplotypes are included in the deeply rooted Mesoamerican-Andean cluster. Moreover, the only Cuban haplotype in the literature and the remaining 2 A2 haplotypes from the Dominican Republic form even more deeply rooted private branches. Similarly, the only haplogroup C1d sample sequenced from the Dominican Republic forms a private branch with the deepest root in a maximum likelihood tree containing 19 additional C1d haplotypes from Mexico to Brasil plus the CRS. In conclusion, our preliminary results suggest that a substantial proportion of the modern native groups in the Caribbean do not share an Amazonian origin with the language their people spoke in 1492. Furthermore, the position of two Dominican lineages at the earliest split in both their respective trees suggests an early origin that could be revealed by further extant lineage extinctions in Mesoamerica and the Andes or an origin in North America.

Resequencing of Australian Aboriginal mtDNA and Y chromosomes. Y. Xue1, M. Cereno Fernandez2, Y. Chen1, S. McCarthy1, MO. Pollard1, Q. Ayub1, N. Nagle2, P. McAllister1, RJ. Mitchell2, C. Tyler-Smith1 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Department of Genetics, La Trobe University, Melbourne, Australia; 3) Griffith University, Gold Coast, Queensland, Australia.

Modern humans originated in Africa and spread across the rest of the globe 50-70 thousand years ago. The first identified divergence outside Africa was between the ancestors of the Australian Aborigines and some nearby populations on the one hand, and the ancestors of Asians, Europeans and other non-Africans on the other hand. The genetic characterization of this divergence and subsequent events in Australian Aboriginal history before the colonial era remain poorly described. Seven Australian Aboriginal males requested sequencing of their mitochondrial DNA (mtDNA) and Y chromosomes, and we generated high depth Illumina 100 bp paired-end sequence data. The average coverage is ~15-20x for Y chromosome and ~3000x for mtDNA. The mtDNAs of these samples belong to new branches of both mtDNA and Y chromosome haplogroup diversity in these populations. The average coverage is ~15-20x for Y chromosome and ~3000x for mtDNA. The mtDNAs of the samples belong to new branches of both mtDNA and Y chromosome haplogroups are known from Australia, but the many new variants illustrate their diversity and distinct origins. The Y chromosomes belong to haplogroups C(3 individuals), K* (3 individuals), both previously known from Australia, and M (1 individual). This last haplogroup has a restricted geographical distribution in Canada and illustrates a genetic link between Australia and that region. The ~3,000 new Y-SNPs present in these samples are permitting refined estimates of the ancient coalescence times between Australian lineages and those in the rest of the world, and re-examining the hypothesis of most recent expansions in Australia and South Asia, originally based on Y-chromosomal similarities.

The possible role of social selection in the distribution of the "Proto-Mongolian" haplotype in Kazakhs, Kyrgyz, Mongols and other Eurasian populations. D.P. Kharisov1, O. Balanovsky2,3,4,5, I. Bogunova1, I. Tazhiulov1, S. Frolova1, Zh. Isakovav1, A. Nimadava5, I. Zakharov2, O. Balanovsky2,3,4,5 1) Center for Life Sciences, Nazarbayev University (Astana, Kazakhstan); 2) Vavilov Institute of General Genetics RAS (Moscow, Russia); 3) Research Centre of Medical Genetics RAMS (Moscow, Russia); 4) Gumiłowy Eurasian National University (Astana, Kazakhstan); 5) The Komsomolsk-on-Amur State University (Komsomolsk-on-Amur, Russia).

Social factors may be important contributors to reproductive success and determination of the selective survival of individuals. Therefore, social selec- tion and other social factors are important for understanding population structure and its formation. The role of social selection on the distribution and formation of Y-chromosomal gene pool has been studied. There is a strong connection between social selection and birth rates of the descendants, whose fathers had achieved high social status during the expansion of the Mongol Empire and associated historical events. A total of 783 haplotypes, including 687 newly obtained and 96 retrieved from the literature were assigned to the haplogroup C3*-M217 (xM48) based on genotyping 17 Y- chromosomal STR markers. These haplotypes represent 11 populations of Eurasia: Kazakhs, Mongols, Kyrgyz, Telengits, Circassians, Balkar, Temir- goys, Karachai, Evenki, Kizhi and the Tuschtsins. At the result, a major haplotype of this language family includes fifteen subtypes. The frequency of this major haplotype within haplogroup C3* was 16.80% in Kazakhs, 10.13% in Mongols and 2.63% in Kirgiz who are not considered as direct descendants of Genghis Khan. 35.10% of the major haplotype was represented by Kazakh tribe Ashamay-Kersey. 17.02% by the Khalkh Mongols and 7.44% by the Barguts. Therefore, we suppose this major ancestral haplotype to be the "proto-Mongolian haplotype", inherited by Genghis Khan and his descendants. It is important to mention that Temujin belongs to Kiyat-Borjigin tribe that in turn is a branch of the bigger Borjigin tribe, of the Khalkh Mongols. Thus, Genghis Khan might be considered as a carrier rather than founder of the star-cluster haplotype. He and his descendants are the ones who contributed to a positive effect of social selection in the distribution of this haplotype. Other examples are the Bar- guts, who had Genghis Khan’s credit and were granted with a number of privileges, or the Kerey, based on the fact that Temujin had been brought up at the court of the Togrul Khan, belonging to the Kerey tribe.
2046T
Juxtapositions of short IBD blocks can cause biased estimation in inferences based on the length of IBD blocks. C.W.K. Chiang1, J. Novembre2. 1) Ecology and Evolutionary Biology, UCLA, Los Angeles, CA; 2) Human Genetics, University of Chicago, Chicago, IL.

Blocks of identity-by-descent (IBD) play an important role in many modern genetic applications, including long-range phasing, imputation, genetic mapping, detection of natural selection, and demographic inferences. One commonly used definition of IBD blocks is that they are contiguous segments of the genome inherited from a recent shared common ancestor without intervening recombination. With programs like Beagle’s fastibd, long IBD blocks (>1cM) can be efficiently detected using high-density SNP array data of a population sample. However, all programs detect IBD based on contiguous segments of identity-by-state (IBS). As such, detected IBD blocks could often be due to the juxtaposition of smaller IBD blocks inherited from different common ancestors. Here, we show the juxtaposition of small IBD blocks leads to an error in estimating the length distribution of IBD blocks and can affect downstream inferences. To demonstrate the prevalence of subsegment juxtapositions, we used coalescent simulations where we know the precise genealogy of the sample and found that >35% of the detected IBD segments of 1cM or longer are composed of at least two subsegments. In particular, 11% of the detectable segments consist of at least 1 other subsegment >25% of the total length, and this effect was more pronounced for detectable segments between 1 to 2cM long, compared to segments >2cM long. To demonstrate that the juxtaposition can lead to practical problems, we investigated the impact on a novel estimator of the de novo mutation rate using IBD blocks. We observed accurate estimates of the input mutation rate when true IBD blocks are used, but overestimates of the mutation rate by ~15 fold using inferred IBD blocks. When the effect of juxtaposition on the estimated age of the block was modeled, the mutation rate estimate improved greatly. Our results suggest that identifying IBD blocks based on extended IBS can inflate the length of IBD blocks, and in this case results in an inflated estimate of the de novo mutation rate, unless properly accounted for. This effect should be carefully considered as methods to detect shorter IBD blocks using sequencing data are being developed.

2047F
Improved detection of ancient hominin admixture in modern humans. S.R. Browning1, B. Verno2, B.L. Browning1,2,3, J.M. Akey1. 1) Biostatistics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA; 3) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA.

The S* statistic [1] detects genomic segments with variants in unusually strong linkage disequilibrium (LD) indicative of ancient admixture with now-extinct hominins such as Neandertals. The detected introgressed haplotypes can be used to reconstruct archaic genomes which will facilitate the inference of important population genetics characteristics such as the effective population sizes of archaic populations, timing and parameters of introgression, and phylogenetic relationships of archaic groups.

S* looks for large LD between a subset of SNPs in a region, with penalties incurred when small numbers of haplotypes enter or leave the introgressed state [1]. Recombination events in humans since admixture are inherited which results in groups of haplotypes entering or leaving the introgressed state together. Instead of counting haplotypes, we count haplotype clusters in the Beagle model [2]. The haplotype cluster approach is particularly advantageous when the number of individuals being analyzed is large, because a large number of haplotypes entering/leaving the introgressed state together is not uncommon. Such haplotype clusters typically can be represented by one haplotype cluster. A single recombination event should be penalized lightly, whereas multiple recombination events should be penalized heavily in the statistic.

To evaluate our method, we simulated sequence data using the model of [3]. In the simulated data, we find that haplotype clusters correlate closely with introgression status. Each haplotype cluster typically represents only introgressed or only non-introgressed haplotypes, even when haplotype phase information is not known a priori. In simulated data with 100 African reference samples and 500 European samples, we find 20% more introgressed regions using our haplotype clustering approach than with the standard S* approach, for the same false discovery rate. Incorporation of haplotype clustering into an S*-type statistic is useful for exploiting the increasing numbers of sequenced human genomes to find introgressed haplotypes.


2048W
The theoretical accuracy of deterministic approximations to coalescent formulas. E.M. Jewett, N.A. Rosenberg. Biology, Stanford University, Stanford, CA.

Many coalescent distributions and expectations can be derived by conditioning on the number, n_i, of alleles at time t in the past that are ancestral to a data set of n t alleles sampled in the present. However, summing over the conditional distribution of n_i, given n_t, can be computationally challenging when n_t is large. Thus, such formulas can be difficult to evaluate on modern genomic datasets with hundreds or thousands of sampled lineages. One alternative to conditioning on all possible values of n_i is to use an approximation in which n_i is assumed to equal its expected value E[n_i] with probability one (Slatkin, 2000). This approximation greatly reduces the number of terms in conditional expressions, significantly reducing their computational complexity. However, despite the utility of the approximation, its theoretical accuracy is not known. Instead, the accuracy of any given version of the approximation must be evaluated empirically by comparing it with the true distribution. As a result, the accuracy of a given approximation cannot be known outside the range of parameter values over which the true distribution can be computed. Here, we show that approximate distributions converge uniformly to the true distributions under certain simple assumptions, and we derive an expression for the asymptotic approximation error. Our results provide a theoretical basis for understanding the ranges of parameter values over which any given approximation is accurate, facilitating the application of the approximation n_t = E[n_t] to reduce the complexity of computing coalescent formulas on large genomic data sets.

2049T
DNA-based detection of Glucose 6-Phosphate Dehydrogenase (G6PD) deficiency alleles in an Eastern Caribbean population. C. Gupta, L. Deschênes, C. Headland, I. McIntosh. American University of the Caribbean School of Medicine, Cupeyco, St. Maarten.

G6PD deficiency is the most common enzyme deficiency worldwide (~400 million cases) with highest prevalence in populations where malaria is endemic. G6PD is required to produce NADPH and maintain glutathione in the reduced state in red blood cells, and deficiency predisposes to haemolytic anaemia and neonatal jaundice in hemizygous males. The frequencies of deficiency alleles have been reported as 0.05 ~ 0.25 in African populations but have not been assessed in the Eastern Caribbean. We developed a noninvasive method to detect the major deficiency alleles through extraction of DNA from buccal cells, PCR amplification and identification of the deficiency alleles via restriction endonuclease digestion and gel electrophoresis. In a cohort self-reporting as African or Afro-Caribbean we identified two hemizygous males and two heterozygous females giving an overall G6PD deficiency allele frequency of 0.048. This result is at the lower end of the range reported previously for which we propose two possible explanations: (1) admixture in the population introducing European & Amerindian alleles (of note, the incidence of Hb-S and Hb-C alleles in this cohort was 0.086 c.f. 0.13 in previous studies of this population); (2) mutation-specific analysis would not detect other deficiency alleles which may be more common in this population.
2050F
Insights on the evolutionary history of Tibetans from whole-genome sequence data. H. Hu1, T. Simonson2, G. Glusman3, J. Roach4, G. Cavalli-Sforza5, M. Brunkow6, M. McCormack7, N. Petousi8, P. Lorenzo9, R. Gelinas2, L. Jorde1, J. Prchal10, P. Robbins10, C. Huff11. 1) Department of Epidemiology, UT MD Anderson Cancer Center, Houston, TX; 2) University of California, San Diego La Jolla, Division of Physiology, San Diego, CA; 3) Institute for Systems Biology, Family Genomics Group, Seattle, WA; 4) Royal College of Surgeons in Ireland, Molecular and Cellular Therapeutics, St Stephen's Green, Dublin 2, Ireland; 5) University of Oxford, Department of Physiology, Anatomy and Genetics, Oxford, OX1 3PT, United Kingdom; 6) University of Utah, Hematology, Salt Lake City, UT; 7) University of Utah, Department of Human Genetics, Salt Lake City, UT.

The Tibetan people inhabit the world's highest plateau, with an average altitude exceeding 4,500 meters. Over thousands of years Tibetans have evolved unique adaptations to this extreme environment, such as low oxygen levels. Here we report the first study of Tibetan evolutionary history from whole-genome sequence data. Using high-coverage whole-genome sequence data from 17 Tibetan and 9 Han Chinese individuals, we characterize recent admixture events and ancient demographic history of Tibetans from SNV allele frequency spectra data in the context of the Out of Africa and European/Asian divergence events. We restrict this analysis to genomic regions at least 10 kb away from exons or conserved regions to minimize the confounding effects of natural selection. Our results indicate that the Tibetan population diverged from the Han Chinese population much earlier than previously estimated from exome sequence analysis. We also detect evidence of an ancient admixture event from Han Chinese to Tibetans around 25,000 years ago, as well as substantial subsequent gene flow between Tibetan and Han Chinese populations. We utilize this demographic model to detect genomic signals of recent positive selection using the Composite of Multiple Signals (CMS) test. The CMS test identifies several genes previously implicated in Tibetan high-altitude adaptation, including EGLN1 and EPAS1. Our results provide a high-resolution map of the potential functional targets of recent positive selection in Tibetan genomes.

2052T
Population structure and selection pressures: short indels and structural variants from NGS in 250 Dutch trios. A. Abdellaiou1, V. Guryev2, F. van Dijk3, P. S.Ramos4, P. Robins1, J. de Ligt5, N. Amin1, F. van Dijk1,12, L. Karssen12, H. Mei6, E.E. Eichler7, D.I. Boomsma1, K. Ye8, Genome of the Netherlands Consortium. 1) Biological Psychology, Vrije Universiteit (VU) Amsterdam, Amsterdam, Noord Holland, The Netherlands; 2) European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 3) Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 6) Life Sciences Group Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands; 7) Section Molecular Epidemiology, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands; 8) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; 9) Department of Genome Sciences, University of Washington, Seattle, Washington, 10) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 11) Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands; 12) Genomics Coordination Center, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands; 13) Netherlands Bioinformatics Centre, Nijmegen, The Netherlands.

While microarray data have contributed much to population genetics, the higher resolution of whole-genome sequence data is expected to yield novel insights. Here we evaluate population structure and selection pressures, and the identification of functional variants under selection. Ancestry differences in the Netherlands show clear geographic distributions, as previously mapped using principal component analysis (PCA) on whole-genome sequencing data. Ancestry-informative PCs can reveal the consequences of selection pressures, and can also be used to detect selection pressures and traces of migration (Abdellaiou et al, 2013). The Genome of the Netherlands (GoNL: http://www.nigenome.nl/) is a whole-genome sequencing project in a representative sample consisting of 250 family trios from all provinces in the Netherlands. Sequencing was done on the Illumina HiSeq 2000 platform at the Beijing Genomics Institute (BGI) on blood-derived DNA from uncultured cells and accomplished coverage was 14x-15x. Several computational pipelines using SNP read depth, copy number variation (CNV), read depth, de novo assembly, split-read alignment, and combined approaches were applied for detection of short indels and different types and size ranges of structural variants (SVs), including deletions, large insertions, tandem duplications, inversions, mobile element insertions and translocations. PCA on short common indels revealed two PCs that show regional differences that were not previously observed, in addition to two PCs showing regional differences previously captured with microarray SNPs (differentially between North and South, and between East and West). Functional variants under selection (and thus with important phenotypic consequences) can be identified by comparing the distribution of alleles between the subpopulations identified by the PCs. Similar analyses are currently being undertaken for additional types of rare and common SVs as well as a combined set of short indels and SVs. [Reference: Abdellaiou et al: Population structure, migration, and diversifying selection in the Netherlands. Eur J Hum Genet 2013; e-pub ahead of print 27 March 2013.]

2051W

PURPOSE: Many common diseases are more prevalent in specific ethnic groups. Given the growing number of disease-associated loci in regions that show evidence of selection, identification of alleles under selection may provide insight into disease susceptibility. Relative to other African-Americans (AA), the AA Gullah population has lower European admixture and higher ancestral homogeneity from the Sierra Leone area in West-West Africa. We sought to capitalize upon the relative closeness between the Gullah and Sierra Leonese to identify regions that differentiate both populations and may hence be under recent population-specific selective pressures.

METHODS: We computed a linear regression model of the HapMap YRI principal component (PC2) as a quantitative outcome, using 777 Gullah and 400 Sierra Leonian samples genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0. We adjusted for European admixture via inclusion of the PC1 component (PC1) as a covariate. In total, 679,513 SNPs with MAF>5% were used in this analysis. In order to exclude spurious loci, only regions where at least one SNP met genome-wide significance (P<5E-10) and a second significant SNP (P<1E-07) in LD with it were considered.

RESULTS: Nine regions met our criteria as those that best differentiate the Gullah from the Sierra Leonian. The most significant was a ~2 Mb region at Xq22.2-q22.3 around the IL1RAPL2 gene, where 4 SNPs had P<5E-10. Other regions included 3q12.3, 4q35.1, 8p23, the extended HLA at 6p22.1-21.32, 7p15.3, 10q11.22, 10q25.1, and 14q24.2. Four additional regions with at least 3 SNPs in LD with P<1E-07 within 1 Mb were identified. Consistent with the literature, our scan identified the HLA and other regions harboring multiple immune-related genes. CONCLUSION: We have identified several regions that differentiate the Gullah from the Sierra Leonians, suggesting that recent selection may be operating at these loci. Given the relative homogeneity of the Gullah and their genetic proximity to Africans from Sierra Leone, identification of regions that might be under selection in the Gullah has the potential to elucidate disease risks in AA.
2053F
1) Department of Cellular and Environmental Biology, University of Perugia, Perugia, Italy; 2) Sorenson Molecular Genealogy Foundation, Salt Lake City, UT 84115, USA; 3) Dipartimento di Biologia e Biotecnologie, Università di Pavia, Italy; 4) Department of Anthropology and Institute for Genomic Biology, University of Illinois Urbana-Champaign, USA; 5) Department of Biological Sciences, Florida International University, Miami, Florida 33199, USA; 6) AncestryDNA, Provo, USA; 7) Département de Pédiatrie, Centre de Recherche du CHU Sainte-Justine, Université de Montréal, Montréal, Canada; 8) Department of Anthropology, University of California, USA; 9) Canadian Museum of Civilization, Gatineau, Canada; 10) Institute for Genomic Biology, University of Illinois Urbana-Champaign, IL 61801, USA.

The old tripartite linguistic subdivision of Native Americans into three major groups has been revived by recently reported patterns of nuclear genome diversity indicating that Native Americans descend from three streams of Asian/Beringian gene flow. Although this scenario was suggested by early mitochondrial DNA (mtDNA) data, neither this model nor the alternative scenarios have been fully evaluated by employing entire mitochondrial genomes. In this study we focused our attention on two North American mtDNA haplogroups, known as A2a and B2a. Analyses of B2a mitogenomes, which are absent in Eskimo-Aleut and northern Na-Dene-speakers, revealed that this haplogroup arose in North America ~11-13 thousand years (ka) ago from one of the founder Paleo-Indian B2 mitogenomes. In contrast, haplogroup A2a, which is typically found among Eskimo-Aleuts and Na-Dene, but also present in the easternmost Siberian groups, originated only 4-7 ka ago in Alaska, led to the first Paleo-Eskimo settlement of northern Canada and Greenland, and contributed to the formation of the Na-Dene gene pool. However, mitogenomes also show that Amerindians from northern North America, without distinction between Na-Dene and non-Na-Dene, were heavily affected by an additional and distinctive Beringian genetic input. In conclusion, most of the contemporary mtDNA variation observed along the double-continental stems from the first wave that from Beringia followed the Pacific coastal route and is dated to ~15-18 ka ago based on the 16 mitogenome founders identified so far. This was accompanied or followed by a second inland migratory event, marked by haplogroups X2a and C4c, which affected all Amerindian groups of Northern North America. Considerably later, the ancestral A2a carriers spread from Alaska undertaking both a westward migration to Asia and an eastward expansion into the circumpolar regions of Canada. Thus, the First American founders left the greatest genetic mark but the original maternal makeup of North American Natives was subsequently reshaped by additional inputs and local population dynamics, making even a three-wave view too simplistic.

2054W
Data from extended 1000 Genomes phase I populations refine comparisons of X-linked and autosomal population genetic patterns. L. Arbiza1, S. Gottipati1, A. Siepel1, A.G. Clark2, A. Keinan1.
1) Dept. of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Dept. of Molecular Biology and Genetics, Cornell University, Ithaca, NY.

Contrasting the patterns of variability on human chromosome X and the autosomes has grown in power and informativeness with the rapid expansion of whole-genome sequence data. Such a comparison can shed light on demographic processes, differences in the histories of males and females, and the past action of natural selection. Previous studies were based on either genotyping data sets, a small set of resequenced loci in six populations, or genomic sequencing data from at most two populations. These studies leave several open questions. Most notably, to what extent do genome-wide patterns generalize to additional populations? Previous studies also differed in their approaches and population genetic measures, which has led to apparent discrepancies among results. For example, estimates of population differentiation (FST) reflect more recent patterns of variation –since the time of split of the specific populations compared– than nucleotide diversity (π), which captures more ancient epochs. To address these questions and extend our previous results, we capitalized on the 1000 Genomes phase I data, consisting of 14 populations from four ancestry-based continental groups. We improved on our methodology for normalizing X-linked and autosomal π by genetic divergence from different outgroups. To obtain refined resolution of different epochs and disentangle previous results, we also estimated FST between each pair of populations and compared the level of differentiation of X-linked and autosomal SNPs along each branch in the population tree. These analyses point to three results: (1) While less sensitive to more recent demographic events than estimates of population differentiation, estimates of π support a reduction of the ratio of X-linked to autosomal effective population size (X/A) in non-African compared to African populations. (2) A wide array of models of human demographic history compiled from the literature only partially predict the magnitude of this observed X/A reduction, supporting the need to invoke additional factors such as sex-biased events during the out-of-Africa dispersal. (3) FST results provide further evidence and accentuate these observations, additionally pointing to variability across different branches, and suggesting changes in mating patterns through human history. Finally, beyond demographic inference, we characterize the relative effect of selection on X and A in different epochs by studying FST as a function of distance from genes.
2055T Population differentiation of two Brazilian populations inferred from non-genic Alu insertions and microsatellites of the Class I Major histo-
compatibility complex. A.C. Aragão1, H.A. S. Botta1, L. Castelli1, E.C. Castelli1, H.R. Magaldi1, J.A. Peña1, S.F. Oliveira1,2,3. 1) Biologia Animal Post-Graduate Program, Institute of Biology, University of Brasilia, Brasilia, Distrito Federal, Brazil; 2) Department of Pathology, School of Medicine of Botucatu, University of the State of Sao Paulo, Botucatu, Sao Paulo, Brazil; 3) Department of Genetics and Morphology, Institute of Biology, University of Brasilia, Brasilia, Distrito Federal, Brazil; 4) Department of Genetics, Physical Anthropology and Animal Physiology, Faculty of Science and Tech-
nology, Universidade Federal do Ceará, Fortaleza, Brazil; 5) Biological Molecular Post-Graduate Program, Institute of Biology, University of Brasilia, Bra-
silia, Distrito Federal, Brazil.

Genetic markers located in important genomic regions, such as genes or regulatory sequences, are most commonly affected by natural selection and thereby behave very differently from neutral regions of the human genome. The human MHC (6p23.1) is a region of great importance due to the presence of genes related to action and regulation of the human immune system. Several studies have reported that natural selection might be acting upon these genes, but few have reported the behavior of non-genic elements in such region. Aiming to infer whether this region is under selection, and not solely its genes, this study intended to evaluate the behavior of the Class I MHC region based on Alu insertions and STR markers located nearby but not within the genes in two Brazilian populations. These populations have been shown to be very different regarding neutral markers of the genome. 122 samples from two distinct populations, an urban admixed population (Brasilia, 65 samples) and an afro-derived rural population (Kalunga, 57 individuals) were studied for their AluMICB, AluTF, AluHJ, AluHG and AluHF microsatellite polymorphisms and the insertion/deletion polymorphisms of Cp1 and Cp2 (10 samples each) from the studied population (200). Preliminary results show a high level of NRY variation and the presence of a high frequency of haplogroup R1a (20%), while Gardenvielf differences from the other communities with a high frequency of haplogroup I2a (33.4%). Overall, the Y-STR profiles of these groups support the documented move-
ments of paternal and maternal prehistories.

2056F Y-chromosomal sequences from southern Africa allow direct compari-
son of paternal and maternal prehistories. C. Barbier1, S. Lippold1, R. Schröder1, S.W. Mpoloka2, M. Stoneking1, B. Pakendorf1, 1) Evolutionary Genetics Institute for Evolutionary Biology, University of Leipzig, Germany; 2) Department of Biological Sciences, University of Botswana, Gaborone, Botswana; 3) Laboratoire Dynamique du Langage, UMR5956, CNRS and Université Lyon Lumiére 2, Lyon, France.

Comparisons of Y-chromosome and non-recombinating Y-chromosome (NRY) variation can be quite informative about sex-biased differences in human demographic history such as population size variation, unbalanced gene flow after contact, and different origins and migration routes; yet NRY analyses are hampered by the common use of ascertained haplogroup defining SNPs, which precludes the possibility to perform many demographic analyses of NRY variation. Recently we developed a capture-based method for enriching Y-chromosome sequence data generated for other African populations, which allows direct comparisons to complete mtDNA genome sequences. We applied this method to a broad DNA collection of ~300 southern African samples predominantly centered on the so-called Khoisan, which was previously analyzed for complete mtDNA genome sequences as well as for autosomal SNPs. Preliminary results show a high level of NRY variation and the presence of highly divergent haplotypes. Our dataset will be compared to the complete mtDNA genome sequences from the same dataset as well as to comparable Y chromosome sequence data generated for other African populations from the CEPH panel, to frame the variability of Khoisan groups within the continent. In particular, we will investigate the extent of sex-biased admixture among groups, the potential contribution of East African pastoralists to southern African groups, and the comparative demographic history of maternal and paternal lineages in southern Africa.

2057W Population structure in five Mennonite communities. K.G. Beatt1, P.E. Melton1, M.J. Mosher1, M.H. Crawford1, 1) Department of Anthropology, University of Kansas, Lawrence, KS; 2) Center for Genetic Origins of Health and Disease, University of Western Australia, Perth, Australia; 3) Department of Anthropology, Western Washington University, Bellingham, WA.

Mennonites are a branch of the Anabaptist religious groups that formed in Europe during the time of the Reformation. Those who followed Menno Simons, the Dutch-North German branch, were later known as Mennon-
ites, and this group later fled from the Netherlands and Germany during the 16th Century to Prussia, and in the 1850s to the United States. The movement and genealogical history of this group has been well documented and indicate population fissions along familial lines after arrival in the United States. This project examines Y-STR profiles consisting of 17 loci from 94 males from the Kansas Mennonite communities of Alexanderwohl (n=13), Lonetree (n=20), Gardenvielf (n=15), and Meridian (n=25), as well as a community of Mennonites from Henderson, Nebraska (n=24). The most prevalent haplotype in all communities was R1b, which represented 56% of the lineages in the entire sample, with other haplogroups represented in frequencies below 10% including E1b, G2a, I1, I2, J2, L, Q, and R1a. There are no shared Y-STR haplotypes between the communities and haplogroup frequencies vary considerably between populations, supporting the historical evidence that original migrating groups split along familial, and specifically, paternal lines upon arrival to the United States. The Alexanderwohl and Henderson communities, which split off from group that migrated from the Ukraine in 1874, are the only two communities with Haplogroups G2a and J2a1. The Meridian community, a heterogenous group believed to be an amalgamation of Swiss, German, and Dutch Mennonites as well as Amish, is represented by 8 haplogroups, including haplogroups F found only in Meri-
dian and a high frequency of E1b1b (12%) when compared to the other populations. It also has a mean number of pairwise differences of 9.3090 (+/- 4.2857) and gene diversity of 0.9621 (+/- 0.0007), higher than any other Mennonite community. Consequently, this group at 65% of the studied individuals, has a high frequency of haplogroup R1a (20%), while Gardenvielf differs from the other communities with a high frequency of haplogroup I2a (33.4%). Overall, the Y-STR profiles of these groups support the documented move-
ments of paternal and maternal prehistories.
Synthesizing genetic and genealogical data to trace historical waves of European and African immigration to the United States. J.K. Bynoe, J.M. Granka 1, K. Noto 1, R.E. Curtis 2, Y. Wang 1, M.J. Barber 1, N.M. Myrnes 1, C.A. Ball 1, K.G. Chahine 1 1) Ancestry.com DNA LLC, 153 Townsend St., Ste. 800, San Francisco, CA; 2) Ancestry.com DNA LLC, 360 West 4800 North, Provo, UT.

Census data can reveal waves of immigration, with immigrants from a single source population migrating to roughly the same U.S. locations within a short period of time. A refinement of the timings and locations of migratory waves using genetic and genealogical data would be useful not only for understanding the peopling of the U.S., but also for understanding patterns of genetic disease risk. To examine migratory waves using genealogical and genetic data, we first identify ‘IBD-enriched groups’: groups of U.S. AncestryDNA customers who share elevated amounts of DNA identity-by-descent (IBD) with a reference panel of individuals with known deep ancestry from a single location. Elevated IBD with individuals from one location could indicate a source population from which ancestors of a customer originated. As a proof of principle, for individuals in a group that is IBD-enriched for a particular location, we search for an enrichment of ancestors born in this location across the collection of their pedigrees. Then, for each IBD-enriched group, we search the collection of pedigrees for enrichment of more recent birth locations, representing migratory destinations within the U.S. Among over 100,000 U.S. AncestryDNA customer samples, we identify IBD-enriched groups who have elevated IBD with more than 20 European countries and seven West African population groups. For each IBD-enriched group, we present maps showing patterns of ancestor birth location enrichment through time. As expected, customers in an IBD-enriched group to a particular location have an enrichment of ancestor births in that location, though this enrichment is dependent on pedigree completeness (e.g. it is rare for African Americans to have known ancestors from West Africa in their pedigrees). For some punctuated migratory waves, such as the arrival of Norwegian immigrants in the Midwest during the 19th century, clear signals of both source and migratory destinations are visible. These locations are temporally ordered, with ancestors from the migratory destination appearing more recently in the pedigrees than ancestors from the source location.

Finally, we use IBD-enriched groups to test regional haplotypes defining IBD-enriched groups. With our approach, we demonstrate that we can estimate individual ancestral origins and detail human migratory history by jointly studying estimates of genetic relatedness along with genealogical data.

The Saudi Arabian Genome Reveals a Two Step Out-of-Africa Migration. J.J. Farrell 1, A.K. Al-Ali 2, L.A. Farrer 1, A.N. Al-Nafaie 1, A.M. Al-Rubaish 1, E. Melota 1, Z. Naserullah 1, A. Alsaleh 1, P. Sebastian 1, M.H. Steinberg 1, C.T. Baldwin 1 1) Department of Medicine, Boston University School of Medicine, Boston, MA; 2) Department of Clinical Biochemistry, College of Medicine, University of Dammam, Dammam, Kingdom of Saudi Arabia; 3) Department of Hematology, College of Medicine, University of Dammam, Dammam, Kingdom of Saudi Arabia; 4) Department of Internal Medicine, College of Medicine, University of Dammam Dammam, Kingdom of Saudi Arabia; 5) Department of Pediatrics, Maternity & Child Hospital, Dammam, Kingdom of Saudi Arabia; 6) Department of Hematology, King Fahd Hospital, Hafof, Al-Ahsa, Kingdom of Saudi Arabia; 7) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Here we present the first high-coverage whole genome sequences from a Middle Eastern population consisting of 14 Eastern Province Saudi Arabians. Genomes from this region are of interest to further answer questions regarding ‘Out-of-Africa’ human migration. Applying a pairwise sequentially Markovian coalescent model (PSMC), we inferred the history of population sizes between 10,000 years and 1,000,000 years before present (YPB) for the Saudi genomes and an additional 11 high-coverage whole genome sequences from Africa, Asia and Europe.

The model estimated the initial separation from Africans at approximately 110,000 YBP. This intermediate population then underwent a long period of decreasing population size culminating in a bottleneck 50,000 YBP followed by an expansion into Asia and Europe. The split and subsequent bottleneck were thus two distinct events separated by a long intermediate period of genetic drift in the Middle East. The two most frequent mitochondria haplogroups (30% each) were the Middle Eastern U7a and the African L. The presence of the L haplogroup common in Africa was unexpected given the clustering of the Saudis with Europeans in the phylogenetic tree and suggests some recent African admixture. To examine this further, we performed formal tests for history of admixture and found no evidence of recent African admixture in the Saudi after the split. Taken together, these analyses suggest that the L3 haplogroup found in the Saudi were present before the bottleneck 50,000 YBP. Given the TMRCA estimates for the L3 haplogroup (400,000 YBP), coupled with the timing of the Out-of-Africa split, these analyses suggest that L3 haplogroup arose in the Middle East with a subsequent back migration and expansion into Africa over the Horn-of-Africa during the lower sea levels found during the glacial period bottleneck.

These results are consistent with the hypothesis that modern humans populated the Middle East before a split 110,000 YBP, underwent genetic drift for 60,000 years before expanding to Asia and Europe as well as back-migration into Africa. Examination of genetic variants discovered by Saudi whole genome sequencing in ancestral African populations and European/Asian populations will contribute to the understanding human migration patterns and the origin of genetic variation in modern humans.

Synthesizing genetic and genealogical data to trace historical waves of European and African immigration to the United States. J.K. Bynoe, J.M. Granka 1, K. Noto 1, R.E. Curtis 2, Y. Wang 1, M.J. Barber 1, N.M. Myrnes 1, C.A. Ball 1, K.G. Chahine 1 1) Ancestry.com DNA LLC, 153 Townsend St., Ste. 800, San Francisco, CA; 2) Ancestry.com DNA LLC, 360 West 4800 North, Provo, UT.

Census data can reveal waves of immigration, with immigrants from a single source population migrating to roughly the same U.S. locations within a short period of time. A refinement of the timings and locations of migratory waves using genetic and genealogical data would be useful not only for understanding the peopling of the U.S., but also for understanding patterns of genetic disease risk. To examine migratory waves using genealogical and genetic data, we first identify ‘IBD-enriched groups’: groups of U.S. AncestryDNA customers who share elevated amounts of DNA identity-by-descent (IBD) with a reference panel of individuals with known deep ancestry from a single location. Elevated IBD with individuals from one location could indicate a source population from which ancestors of a customer originated. As a proof of principle, for individuals in a group that is IBD-enriched for a particular location, we search for an enrichment of ancestors born in this location across the collection of their pedigrees. Then, for each IBD-enriched group, we search the collection of pedigrees for enrichment of more recent birth locations, representing migratory destinations within the U.S. Among over 100,000 U.S. AncestryDNA customer samples, we identify IBD-enriched groups who have elevated IBD with more than 20 European countries and seven West African population groups. For each IBD-enriched group, we present maps showing patterns of ancestor birth location enrichment through time. As expected, customers in an IBD-enriched group to a particular location have an enrichment of ancestor births in that location, though this enrichment is dependent on pedigree completeness (e.g. it is rare for African Americans to have known ancestors from West Africa in their pedigrees). For some punctuated migratory waves, such as the arrival of Norwegian immigrants in the Midwest during the 19th century, clear signals of both source and migratory destinations are visible. These locations are temporally ordered, with ancestors from the migratory destination appearing more recently in the pedigrees than ancestors from the source location.

Finally, we use IBD-enriched groups to test regional haplotypes defining IBD-enriched groups. With our approach, we demonstrate that we can estimate individual ancestral origins and detail human migratory history by jointly studying estimates of genetic relatedness along with genealogical data.

The Saudi Arabian Genome Reveals a Two Step Out-of-Africa Migration. J.J. Farrell 1, A.K. Al-Ali 2, L.A. Farrer 1, A.N. Al-Nafaie 1, A.M. Al-Rubaish 1, E. Melota 1, Z. Naserullah 1, A. Alsaleh 1, P. Sebastian 1, M.H. Steinberg 1, C.T. Baldwin 1 1) Department of Medicine, Boston University School of Medicine, Boston, MA; 2) Department of Clinical Biochemistry, College of Medicine, University of Dammam, Dammam, Kingdom of Saudi Arabia; 3) Department of Hematology, College of Medicine, University of Dammam, Dammam, Kingdom of Saudi Arabia; 4) Department of Internal Medicine, College of Medicine, University of Dammam Dammam, Kingdom of Saudi Arabia; 5) Department of Pediatrics, Maternity & Child Hospital, Dammam, Kingdom of Saudi Arabia; 6) Department of Hematology, King Fahd Hospital, Hafof, Al-Ahsa, Kingdom of Saudi Arabia; 7) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Here we present the first high-coverage whole genome sequences from a Middle Eastern population consisting of 14 Eastern Province Saudi Arabians. Genomes from this region are of interest to further answer questions regarding ‘Out-of-Africa’ human migration. Applying a pairwise sequentially Markovian coalescent model (PSMC), we inferred the history of population sizes between 10,000 years and 1,000,000 years before present (YPB) for the Saudi genomes and an additional 11 high-coverage whole genome sequences from Africa, Asia and Europe.

The model estimated the initial separation from Africans at approximately 110,000 YBP. This intermediate population then underwent a long period of decreasing population size culminating in a bottleneck 50,000 YBP followed by an expansion into Asia and Europe. The split and subsequent bottleneck were thus two distinct events separated by a long intermediate period of genetic drift in the Middle East. The two most frequent mitochondria haplogroups (30% each) were the Middle Eastern U7a and the African L. The presence of the L haplogroup common in Africa was unexpected given the clustering of the Saudis with Europeans in the phylogenetic tree and suggests some recent African admixture. To examine this further, we performed formal tests for history of admixture and found no evidence of recent African admixture in the Saudi after the split. Taken together, these analyses suggest that the L3 haplogroup found in the Saudi were present before the bottleneck 50,000 YBP. Given the TMRCA estimates for the L3 haplogroup (400,000 YBP), coupled with the timing of the Out-of-Africa split, these analyses suggest that L3 haplogroup arose in the Middle East with a subsequent back migration and expansion into Africa over the Horn-of-Africa during the lower sea levels found during the glacial period bottleneck.

These results are consistent with the hypothesis that modern humans populated the Middle East before a split 110,000 YBP, underwent genetic drift for 60,000 years before expanding to Asia and Europe as well as back-migration into Africa. Examination of genetic variants discovered by Saudi whole genome sequencing in ancestral African populations and European/Asian populations will contribute to the understanding human migration patterns and the origin of genetic variation in modern humans.

Posters: Evolutionary and Population Genetics
The population history of African and Caribbean vervet monkeys inferred from 130 whole-genome-sequenced samples. Y. Huang1, H. Svarda2, C.A. Schmitt1, J.K. Byrnes2, J.H. Hussin1, M.J. Barber2, N.M. Freimer1, R.K. Wilson7, K. Noto2, J.-C. Grenier1, J.W. Wasserscheid1, N. Juretic1, J.-P. Goulet1, J.-C. Grenier1, M. Capredon1, A. Hodgkinson1, T. de Maller1, V. Brust1, E. Gbeha1, E. HipK1, P. Awadalla1,2, J. Hussin1, J.-P. Goulet1, N. M. Freimer1, J. Hussin1, J. K. Byrnes2, J. W. Wasserscheid1, N. Juretic1, J. C. Grenier1, M. Capredon1, A. Hodgkinson1, T. de Maller1, V. Brust1, E. Gbeha1, E. HipK1, P. Awadalla1,2, 1) Sainte-Justine Hospital Research Center, Université de Montréal, Montreal, Canada; 2) Department of Social and Preventive Medicine, Université de Montréal, Montreal, Canada; 3) Department of Pediatrics, Université de Montréal, Montreal, Canada.

The province of Quebec was colonized four hundred years ago by settlers coming from France. In the early settlement of Quebec, some First Nations populations were allied to French settlers and exchanged genetic material as the result of mixed unions between an aboriginal mother and a French father. After the British Conquest, the rapidly expanding French Canadian population generally grew in relative isolation with limited exchange with British, and other incoming populations such as Acadians and Loyalists. The colonization of the territory took place in successive waves leading to regional founder effects contributing to the uniqueness of the French Canadian population. Today, about 80% of the province’s 8 million inhabitants are French speaking. Starting in 2010, over 30,000 people from the province of Quebec were recruited to be a part of the CARTaGENE project. Genotyping data (Illumina Omni2.5M) was generated for ~1000 participants sampled in three distinct regions of the province: the Montreal area, the Quebec City area and the Saguenay region to use the genotypes to study regional differentiation due to demographic history, admixture and migration patterns within Quebec. We analyze identity by descent (IBD) segments to infer relatedness between participants and trace back regional populations flowing inside Quebec. Inferred IBD segments are also analyzed to attest their ancestral origin. We compare and contrast inferred and reconstructed recombination breakpoints across the genome. We also investigate local ancestry of participants with a panel consisting of Native American and European populations. Specifically, we (i) identify regions of the genome shared between individuals and inherited from common ancestors; (ii) characterize IBD tracts in relation to recombination map and (iii) infer local ancestry across the genome to quantify the relative contributions of the different populations. This work refines previous analysis of population structure in Quebec by providing more accurate demographic histories of descent contributed to the actual French Canadian genetic pool and gives insight on how their contribution is linked to mutations causing differences in disease prevalence throughout regions of Quebec.

Reconstruction of Ancestral Human Genomes from Genome-Wide DNA Matches. J.M. Granka1, R.E. Curtis2, J.K. Byrnes1, M.J. Barber1, N.M. Myres2, K. Noto1, Y. Wang3, C.A. Ball1, K.G. Chahine1. 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Provo, UT.

Individuals who lived long ago may still have much or all of their genome present in modern populations. The genomes of these individuals exist in small segments broken down by recombination and inherited in part by his or her descendents. If such an individual had many children, leading to a large number of descendents today, much of the ancestral genome will be present in modern populations. For the pairs of descendents with the 'target' ancestor as their most recent common ancestor (MRCA), any region of their genomes shared identical-by-descent (IBD) most likely represents the corresponding region of the ancestor's genome. Given a set of pairs of individuals linked to the same MRCA, we develop a novel computational approach to reconstruct the haplotypes of the MRCA from the IBD segments and haploypes of the descendents. With simulated data we assess the performance of our method, affected by factors such as quality of genealogical trees used to infer the MRCA, reliability of inferred IBD, coverage of IBD segments, number of descendents of the MRCA, and number of sampled descendents. To demonstrate the utility of our method, we examine over 125,000 individuals in the AncestryDNA database with phased genome-wide single nucleotide polymorphism data and detailed genealogical information. After first identifying regions of the genome shared IBD between all individuals, we selected one group of several hundred individuals with an 18th century as a known MRCA. Using our method to tile together these individuals’ IBD segments, we are able to reliably construct the ancestral couple’s four haplotypes in large genomic regions with high coverage of IBD segments. In regions of the genome with lower IBD coverage, we are unable to identify and construct all haplotypes with certainty. Our study demonstrates the possibility of reconstructing the genomes of human ancestors, with large family sizes and a large number of living descendents, who lived one to even 12 generations ago. The ability to reconstruct the genomes of human ancestors using genetic and genealogical data has exciting implications in the fields of population genetics, medical genetics, and genealogy research.
2065F The Sherpa and Tibetans share a common genetic history and adaptations to high-altitude. C. Jeong¹, D.B. Witonsky¹, G. Alkorta-Aranburu², B. Basran³, M. Neupane⁴, J.K. Pritchard⁵, C.M. Beall⁶, A. Di Rienzo⁷, 1) Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America; 2) Nepal International Clinic, Kathmandu, Nepal; 3) Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America; 4) Department of Anthropology, Western Research University, Cleveland, Ohio, United States of America.

The Sherpa are an ethnic group in Nepal, traditionally living in high-altitude (HA) regions of the Himalayas. Linguistic and historical evidence indicates a close relationship with Tibetans. Furthermore, the Sherpa share a low concentration of hemoglobin, a key physiological adaptation to HA hypoxia. These findings suggest that the Sherpa and Tibetans constitute closely related branches of the East Asian (EA) populations sharing phenotypic HA adaptations. To develop this idea further, we studied the genetic history of the Sherpa in relation to other Asian populations. First, we asked if the Sherpa are genetically distinct by comparing genome-wide genotyping data of 69 Sherpa native residents at 3800m with the Human Genome Diversity Panel data using principal component analysis (PCA) and ADMIXTURE. These show that the Sherpa genome contains a large portion of a genetic component clearly distinguishable from lowland EA populations. Sherpa individuals have neither South nor Central Asian genetic components and they are tightly clustered with other EA individuals in the global PCA plot. Second, we inferred a past population size change of the Sherpa by applying the pairwise sequential Markovian coalescence method to whole genome sequences (27-30x) of two males. We compared the Sherpa results with those of Han and Dai individuals, which were sequenced at similar depth (Meng et al., 2012). This comparison suggests that the early split of the Sherpa and lowland EA populations about 20,000-30,000 years ago and no evidence for the recent population expansion that characterizes the history of other EA populations. We estimate the effective population size of the Sherpa after the split at about 4,000 individuals. Third, a maximum-likelihood tree of EA populations supports a branch consisting of the Sherpa and Tibetans (from Yunnan Province of China; Beall et al, 2010). Last, we found that 21 of 36 EPAS1 (endothelial PAS domain-containing protein 1) gene SNPs (with a significant association with hemoglobin level in a previous study of Tibetans (Beall et al, 2010) also show an association with nomininal p < 0.05, all of which have the same direction of allelic effect as in Tibetans. The estimated effect sizes also match the Tibetan results. To summarize, our findings support a shared genetic history and genetic adaptation to HA of the Sherpa and Tibetans.

2066W The Genetic Architecture of Skin Pigmentation in Southern Africa. A.R. Martin¹, J.M. Granka², C.R. Gignoux³, M. Möller³, C.J. Wernery³, J.M. Kidd³, M.W. Feldman⁴, E.G. Hoat⁵, C.D. Bustamante¹, B.M. Henn¹, 1) Stanford University, Genetics Department, Stanford, CA; 2) Ancestry.com, San Francisco, CA; 3) University of California, Pharmaceutical Sciences, San Francisco, CA; 4) Stellenbosch University, Health Sciences, Tygerberg, South Africa; 5) University of Michigan, Department of Human Genetics, Ann Arbor, MI; 6) Stanford University, Department of Biological Sciences, Stanford, CA; 7) SONY, Department of Ecology and Evolution, Stony Brook, NY.

Skin pigmentation is one of the most recognizably diverse phenotypes in humans across the globe, but its highly genetic basis has mainly been studied in northern European and Asian populations. The Eurasian pigmentation alleles are among the most differentiated variants in the genome, suggesting strong positive selection for light skin pigmentation. Light skin pigmentation is also observed in the far southern latitudes of Africa, among KhoeSan hunter-gatherers of the Kalahari Desert and other populations. The KhoeSan hunter-gatherers, believed to have diverged from other populations 100,000 years ago, maintain extraordinary levels of genetic diversity, but it is unknown whether light skin pigmentation represents convergent evolution or the ancestral human phenotype. We have collected saliva samples, ethnographic information, and pigmentation phenotypes from 123 individuals in the Kxomani San from the Kalahari. To understand the genetic basis for light skin pigmentation, we have genotyped and exome sequenced 91 Kxomani San individuals to high coverage, generating one of the largest indigenous African exome datasets sequenced outside of 1000 Genomes. Because linkage disequilibrium decay is rapid in this population, we have assessed parameters influencing phasing and imputation accuracy since ideal reference panels do not exist. We have also pursued multiple genotype/phenotype mapping methods, including a mixed model approach, admixture mapping, and linkage mapping. After controlling for admixture from European and Bantu-speaking populations, we find that globally common variants are not significantly associated with pigmentation. Rather, our results indicate that there are a multitude of rare variants in known pigmentation genes, and suggest that previously unidentified genes acting in canonical pigmentation pathways may be involved. Our results highlight the strength of diverse population studies to explain phenotypic variation in the context of human evolutionary history.
Y-chromosomal variation in native South Americans: bright dots on a gray canvas. M. Nothnagel1, 4, 5, L. Roever1, 5, L. Gusmão2, 5, V. Gomes4, M. González4, D. Corach4, A. Sala8, E. Alechines3, F. Paima3, N. Santos3, A. Ribeiro-dos-Santos5, M. Geppert5, S. Willuweit3, M. Nagy5, S. Zwynert5, M. Baeta8, C. Núñez8, B. Martínez-Jarreta8, F. González-Andrade8, E. Fagundes de Carvalho8, 12, D. Aparecida da Silva14, J. José Bulles14, 12, D. Turbón1, 12, A. Maria Lopez Parra14, E. Arroyo-Pardo14, U. Toscanini15, 16, C. Barletta17, S. Santos5, M. Krawczak2, 1) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 2) Institute of Medical Informatics and Statistics, Christian-Albrechts University, Kiel, Germany; 3) Institute of Legal Medicine and Forensic Sciences, Department of Forensic Genetics, Charité - Universitätsmedizin Berlin, Berlin, Germany; 4) Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal; 5) Universidad Federal do Pará, Laboratório de Genética Humana e Médica, Belém, Pará, Brazil; 6) Universidad de Buenos Aires, Facultad de Farmacia y Bioquímica, Servicio de Huellas Digitales Genéticas, Buenos Aires, Argentina; 7) Departamento de Psiquiatría y Psicoterapia, Charité-Universitätsmedizin Berlin, Berlin, Germany; 8) Department of Forensic Medicine, University of Zaragoza, Zaragoza, Spain; 9) Science and Technology Department, Ministry of Public Health, Quito, Ecuador; 10) Laboratorio de Diagnósticos por DNA, Instituto de Biología, Universidad de Estudo do Rio de Janeiro, Rio de Janeiro, Brazil; 11) GENES Ltda, Laboratorio Genética Forense y Huellas Digitales del DNA, Medellín, Colombia; 12) Instituto de Biología, Universidad de Antioquia, Medellín, Colombia; 13) Unidad d’Anthropología, Departamento de Biología Animal, Facultad de Biología, Universitat de Barcelona, Barcelona, Spain; 14) Laboratorio de Genética Forense y Diagnóstico de Enfermedades Infecciosas, Facultad de Medicina, Universidad Complutense, 28040 Madrid, Spain; 15) PRICAI-Fundación Favaloro, Buenos Aires, Argentina; 16) Laboratorio de Genética Molecular, Unidad de Genética Médica, Facultad de Medicina, Universidad del Zulia, Maracaibo, Venezuela; 17) Facultad de Ciencias Biológicas, UNMMS-Universidad Nacional Mayor de San Marcos, Lima, Peru.

While human populations in Europe and Asia have often been reported to reveal a concordance between their extinct genetic structure and the prevailing regional pattern of geography and language, such evidence is lacking for South Americans. In the first such study of South American natives to date, we examined the relationship between Y-chromosomal genotype on the one hand, and male geographic origin and linguistic affiliation on the other. We observed virtually no structure for the extinct Y-chromosomal genetic variation of South American males that could sensibly be related to their inter-tribal geographic and linguistic relationships, augmented by locally confined Y-STR autocorrelation. Analysis of repeatedly taken random subsamples from Europe adhering to the same sampling scheme excluded the possibility that this finding was due to our specific scheme. Furthermore, for the first time, we identified a distinct geographical cluster of Y-SNP lineages C-M217 (C3*) in South America, which are virtually absent from North and Central America, but occur at high frequency in Asia. Our data suggest a late introduction of C3* into South America no more than 6,000 years ago and low levels of migration between the ancestor populations of C3* carrier and non-carriers. Our findings are consistent with a rapid peopling of the continent, followed by long periods of isolation in small groups, and highlight the fact that a pronounced correlation between genetic and geographic/cultural structure can only be expected under very specific conditions.

Analysis of haplotype sharing and recent demographic history in the Netherlands. P. Palamara, T. Pe’er, The Genome of Netherlands Consortium: Columbia University, New York City, NY.

Chromosomal segments that are identical by descent (IBD) were recently shown to convey information about population-level features such as demography, natural selection and heritability of common traits. In a recent work [1], we have developed analytical models for the relationship between haplotype sharing and demography, and shown that IBD sharing provides an effective way for reconstructing demographic events of the recent millennia, where classical methods are typically underpowered. We now extend the developed models to accommodate the simultaneous analysis of multiple demes, providing insight into recent migration rates as well as population size fluctuations. Using this approach we analyzed sequencing data for 498 unrelated individuals from 11 Dutch provinces (the Genome of Netherlands Project). Pairs of individuals from all the analyzed provinces are found to share several IBD segments of length greater than 1 centimorgan (cM), suggesting recent common ancestry of these groups. We observe a north-to-south gradient of declining IBD sharing frequency. While the chance of sharing long (>7 CM), extremely recent IBD segments correlates with modern-day geographic distance, shorter segments are more frequently shared with individuals currently residing in the north of the country, regardless of the individuals’ modern location. Using the developed analytical methods, we reconstruct coalescent distributions and migration rates across the analyzed provinces. In all cases we find evidence for recent exponential growth at different rates for different provinces, with substantial recent gene flow between these demes. Using the retrieved model, we estimate the average haploid pair of Dutch individuals in the studied dataset to find a common ancestor ~1600 years before present, with earlier common ancestors typically found in northern provinces and variation that depends on modern geographic location.


Limitations to determining genetic history in the recombinant genome and connection to demographic events from samples of modern populations. D. Platt, F. Utro, L. Pandit. IBM T. J. Watson Research Center, Yorktown Heights, NY.

We report on our study of the reconstructable genetic history and connections to demographic events impacting genetic history in the recombinant genome derived from modern population samples. We used COSI, which allows for specification of demographic events, such as the out-of-Africa migrations, glacial isolation, and post glacial and agricultural expansions, along with parameterizations for African, European, Asian, and African-American populations, and which allows for the stochastic construction of multi-population histories consistent with these population demographic parameters. In prior work, we constructed Ancestral Recombination Graphs (ARGs) to represent the full genetic history of these multiple simulations of these populations. We mathematically defined and characterized Minimal Descriptor (MD) ARGs capturing the essential substructure of an ARG which preserves the genetic landscape of the extant samples, including the toplogy and edge lengths of the marginal trees. As such, these MDs derived from full COSI output, contain the largest amount of information uniquely specified by extant lineages. We identified a minimal set of the number of nodes in a MD due to recombination and coalescence in the minimal descriptor, and measured the number of nodes within any given epoch that were reconstructable from the entire modern population. We showed that coalescence causes the loss of information defining the recombination history, which information loss can include which genome segments had evolved together. This information loss is similar to that observed for equivalent SNPs in non-recombinant phylogenies, where the order in which the SNPs accumulated is lost. Up to 35% of the reconstructable nodes are lost. Further, the reconstructable histories consistently revealed underlying demographic events, even across radically different histories. Here we report, for a given specific history, the reconstructability of the genetic record from samples drawn from the modern population. We generated a representative set of 1,000 samples with varying sample sizes, seeking to quantify how many of the historical nodes were reconstructable from each of the samples, giving a representation of how much of the history could be expected to be recovered from any given sample. We have found the older ARG nodes are broadly found across multiple recombinants, while more detail in recent events become available with larger sample sizes.
Analyses of exome variants suggest enrichment for functionally-important variants during human expansion into the Americas. S. Ramachandran1,2, J.J. Yang2, 1) Ecology & Evolutionary Biol, Brown Univ, Providence, RI; 2) Center for Computational Molecular Biology, Brown Univ, Providence, RI; 3) Department of Pharmaceutical Sciences, St. Jude Children’s Research Hospital, Memphis, TN.

Genetic variation bears signatures of demographic processes such as population divergences, and a modest number of non-coding variants is adequate to classify human individuals into clusters corresponding to their ancestry. In contrast, the population structure of coding variants has not been well characterized, due to (1) low sampling of non-European individuals for exome- and whole-genome studies and (2) controlling for ancestry in genome-wide association (GWA) studies to avoid spurious correlations. However, controlling for ancestry overlooks the fact that haplotypic and allelic diversity decreases with genetic distance from Africa, meaning derived alleles and variants with functional importance may be particularly enriched on East Asian and Native American ancestral backgrounds. We set out to comprehensively examine the relationships between genetic ancestry and coding variant frequency in admixed US Hispanic populations. In a preliminary analysis of 80 unrelated Mexican individuals from the 1000 Genomes (MXL), we evaluated associations between East Asian, African, Native American, and European ancestry coefficients with the number of copies of the reference allele at polymorphisms on the Illumina Exome chip. African and East Asian ancestry is least often associated (<5%) with number of copies of the reference allele. In contrast, the proportion of SNPs associated with levels of Native American ancestry is 10.3%, slightly lower than those associated with European ancestry (11.2%), and Native American ancestry is significantly correlated with copies of the non-reference allele (and the derived allele) in almost all of these cases. The differences observed in allele frequencies at exome SNPs on different ancestral backgrounds suggest that phenotypically-important variants are enriched on Native American background, which may explain differential incidence in certain diseases – such as acute lymphoblastic leukemia (Xu et al. JNCI 2013) – for individuals with >10% Native American ancestry.

Whole exome sequencing of 126 Northeast Asian individuals including Korean, Chinese, Japanese, and Mongolian ethnicities. A. Rhie1,2,3, S. Lee1,2,3, W. Roh1, J. Shin1, J.J. Kim1,2,3, J.S. Seo1,2,3, 1) Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul 110-799, Korea; 2) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea; 3) Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul 110-799, Korea; 4) Macrogen Inc., Seoul 153-023, Korea.

Along with the development of massively parallel sequencing methods, human DNA sequencing data have been accumulated at a very rapid rate these days. However, most of them were to focus on target diseases or phenotypes, and genomic information of normal individuals has not been well studied so far for each population. Here we provide the sequence information of Northeast Asian exomes, along with some sequence-level and population-level analyses. We sequenced 126 exomes of Asian individuals; 37 Koreans, 25 Chinese, 25 Japanese, and 39 Mongolians. Variations including single nucleotide polymorphism (SNP), insertion/deletion, and copy number variation were called for these samples, and the genotype frequencies of sequenced bases were provided including wild type bases (http://asap.gmi.ac.kr). The bases with low genotype qualities were explored genome-wide, and bases with ≥50% low quality genotypes were found to be largely overlapped with segmental duplications (37.99%), which include several genes such as NBPFT10, FLG, FGR1, MUC6, and TBC1D3C. In addition, we tried to identify genetic diversity and population structure of the Northeast Asians and reconstruct the demographic history of these populations. As a result, we showed evidences of the Northern route migration in the East Asia and suggested some Mongolian-specific functional loci (rs1453544: OR4D6, rs1453547: OR5A2, rs1418138101: ARHGAP3, rs202130413: ARSB, etc.). In this study, we provided the exonic sequence information of the Northeast Asian populations. We hope that our data provide fundamental information to those studying on Asian populations in genetic or medical research fields.

The impact of background selection on fine-scale population structure in humans. R. Torres1,2, RD. Hernandez3,2, 1) Biomedical Sciences Graduate Program; 2) Department of Bioengineering and Therapeutic Sciences.; 3) UCSF, San Francisco, CA.

Patterns of observed genetic polymorphism across human populations have provided great insight about those populations’ respective demographic histories and their underlying population structure. Due to geographic and cultural barriers, reduced gene flow has resulted in different distributions of allele frequencies between human populations through the process of genetic drift. Population genetics methods and measures, such as principle components analysis (PCA) and Fst, have uncovered the genetic structure that exists between human populations, especially at the intercontinental scale. However, because of the processes of natural selection and recombination, we should not expect that the effective population size of an individual will be homogenous across the entire genome. Rather, diversity-reducing selection (e.g., genetic hitchhiking and background selection) near functional loci has resulted in a mosaic of different effective population sizes across the human genome. Population genetics theory predicts that the strength of drift is a function of both effective population size and time. Leveraging this fact, we have demonstrated through PCA that loci contributing to principle components of populations across European populations of the POPRES dataset are strongly correlated with background selection in 7 of the top 10 PCs (Kendall’s tau, significance assessed by permutation). Similar patterns were replicated with respect to Fst as a function of background selection across the same populations. By controlling for conservation, we will present new analysis to demonstrate that such patterns are not the result of direct selection at each site but due to linkage disequilibrium. Thereby we highlight the fine-scale population structure that exists at the intra-continental scale and will illuminate valuable targets for the modeling of subtle population demographic histories.

Evaluating the impact of sex-biased demography and selection on genomic patterns of human diversity. A. Walia1, W. Fu2, S. Ramachandran1,2, 1) Ecology & Evolutionary Biology, Brown University, Providence, RI 02912, USA; 2) Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA; 3) Centre for Computational and Molecular Biology, Brown University, RI 02912 USA.

Disentangling the effects, and relative roles, of demography and selection is central to human population genetics. Due to its model of transmission, the X chromosome is an ideal system with which to gain insight into both natural selection and sex-biased demographic processes in the human lineage. Recessive deleterious mutations on the X are exposed to selection in males, and the smaller effective population size of the X means the X chromosome experiences drift more strongly than do the autosomes. The ratio of the genetic diversity normalized by divergence between the X chromosome and the autosomes ($\rho_{X/\Aut}$) is expected to be 0.75. However, past studies (Hammer et al. 2009, Keinan et al. 2009, Hammer et al. 2010, Gottipati et al. 2011) report equivocal estimates for this ratio, and offer different explanations for observed deviations from $\rho_{X/\Aut}=0.75$. Here, we analyze whole genome data of 567 female individuals from 14 populations, genotyped in Phase 1 of the 1000 Genomes Project. To study the effect of selection on observed genetic diversity, we partition polymorphisms by distance to their nearest gene. We then calculate the ratio of genetic diversity normalized by divergence on the X chromosome to that on the autosomes in different distance intervals to the nearest gene. We compare the $\rho_{X/\Aut}$ ratio for all pairs of 1000 Genomes populations. Gottipati et al. 2011 suggest that male-biased migration out-of-Africa led to a smaller effective population size of females in non-African populations based on their observation of $\rho_{X/\Aut}<1$. Their analysis focused on Utah Residents from the CEPH panel with Northern and Western European ancestry(CEU) and Yoruba in Ibadan, Nigeria(YRI). We analyze all 1000 Genomes Phase 1 populations to identify genomic signatures of sex-biased demographic events and social structures in human evolutionary history. This study allows us to examine the relative impact of locally-acting selection and sex-biased demographic processes on human genome-wide variation.

Effective population size ($N_e$) and population divergence time (TD) are two fundamental parameters to understand demographical history of modern human populations. Taking advantage of genome-wide high density single nucleotide polymorphism (SNP) data, we estimated $N_e$ and TD in 53 populations worldwide based on linkage disequilibrium (LD) information across genome. The estimation by LD-based approach makes use of large number of closely linked markers and makes it possible to show changes of $N_e$ traced over time. Our results revealed the stable African populations and the expansion of non-African populations. We further used two independent approaches to estimate the divergence time between populations, which gave consistent estimation. Interestingly, our estimation of divergence time between European and African was much lower than that between Asian and African, suggesting more gene flow between European and African populations since the initial population divergence. The results in this study will advance our understanding of population demographic history, especially, the “out of Africa” model in the early history of human migration and recent interaction between populations since the initial divergence.

2076T Spatial Localization At Each Locus in the Genome. W. Yang¹, A. Platl², C. Chiang³, E. Eskin⁴, J. Novembre⁵, B. Pasaniuc⁶. ¹) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA; ²) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA; ³) Department of Pathology and Laboratory Medicine, Geffen School of Medicine at University of California, Los Angeles, Los Angeles, CA; ⁴) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA; ⁵) Department of Human Genetics, University of Chicago; ⁶) Department of Pathology and Laboratory Medicine, Geffen School of Medicine at University of California, Los Angeles, Los Angeles, CA.

Ancestry analysis plays an important role in studies of human disease and evolution. Traditionally, ancestry inference from genetic data has been focused on modeling populations as discrete sources. As a result, ancestry prediction at each locus in the genome of any given individual can only be performed in terms of discrete population assignments or ancestry proportions. We have recently proposed methods for ancestry inference in a geographic continuum, by explicitly modeling the spatial distribution of each allele at any locus in the genome as a continuous function of geographic space (Yang et al Nat Genet 2012, Baran et al AJHG 2013). A major drawback of these approaches is that they only model individuals as having ancestry from one or at most two geographical locations (whereas, for example, 20% of the individuals collected in the POPRES data have recent ancestors from at least two country origins). In this work we propose approaches for continuous localization from genetic data for individuals that have recent ancestry from multiple geographical locations. That is, our methods seek to localize on a continuous map the ancestry of the recent ancestors (parents, grandparents and in general all ancestors up to several generations) using the genetic data of a given individual. In addition, we leverage the inheritance pattern to assign geographical locations for each segment in the genome according to the location of its ancestors. We devise an efficient Expectation Maximization approach within hidden Markov models of ancestry in conjunction with forward-backward algorithm that assigns continuous local ancestry for a given individual efficiently (e.g. an average of 2 minutes per individual). We use real data from POPRES as well as admixed individuals of Latino ancestry to show that our method yields high accuracy in ancestor localization as well as in assigning continuous ancestry at each locus in the genome. In simulations starting from POPRES data, our method accurately infers the locations of the 2 grandparents of a given individual’s phased haplotype within an average of around 400Km of the true location and assigns continuous local ancestry within around 400Km of the true location at each segment in the genome. Software package implementing our methods is freely available in our website http://genetics.cs.ucla.edu/spa.

2077F Separation of the largest eigenvalues in eigenanalysis of genotype data from discrete subpopulations. K. Bryc¹, W. Brycª, J.W. Silverstein². ¹) Department of Genetics, Harvard Medical School, Boston, MA; ²) Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH; ³) Department of Mathematics, North Carolina State University, Raleigh, NC.

Principal component analysis (PCA) has been a powerful and efficient method for analyzing large datasets in population genetics since its early applications by Cavalli-Sforza and others. In particular, PCA of single nucleotide polymorphism genotype data can be used to illuminate population structure. We present a mathematical model, and the corresponding mathematical analysis, that justifies and quantifies the use of principal component analysis of biallelic genetic marker data for a set of individuals to detect the number of subpopulations represented in the data. A good estimate for the number of populations, $K$, is needed in Bayesian clustering algorithms such as STRUCTURE (Falush et al. 2003) or ADMIXTURE (Alexander et al. 2009), where one must specify a priori the number of clusters in the data, which affects the inferred relationships among individuals. In contrast to previous work, our results describe behavior of the eigenvalues of the sample covariance matrix without centering or normalization, taking into account both the number of individuals and the number of markers. The raw unprocessed covariance matrix is more amenable to mathematical analysis, and the singular values of such raw data exhibit quantifiable properties that can be used directly to determine the number of populations in the data in an almost deterministic fashion, at least when the number of individuals in the study is sufficiently large. We show that for large data sets of individuals from K well-differentiated subpopulations, with overwhelming probability the un-centered sample covariance matrix has K large eigenvalues. We demonstrate two proof-of-principle simulations that we are able to obtain evidence of population structure when the number of individuals is large enough. We indicate that the power of the technique relies more on the number of individuals genotyped than on the number of markers.


The distribution of genomic regions that are shared Identical by Descent (IBD) among individuals can provide insight into population history, facilitate the identification of adaptively evolving loci, and is an important tool in disease gene mapping. Next-generation sequencing technology has enabled the creation of large exome sequencing datasets from thousands of individuals. However, accurately inferring IBD segments from exome data is difficult because of the sparsity of the data. Here, we describe an accurate and robust computational framework to detect IBD in exome data, which we evaluated through comprehensive simulations. The key insight of our approach is to identify and exclude exonic loci that are refractory to accurate IBD calling because of insufficient exon density. We applied this computational framework to detect IBD segments in high-coverage exome sequences from 4,298 European-American individuals studied as part of the NHLBI Exome Sequencing Project. Patterns of IBD sharing reveal significant evidence for cryptic and spatial population structure, and allow us to infer the average number of meioses that separate two randomly selected European-American individuals. Overall, our results enable the power of IBD analyses to be applied to exome data and reveal novel insights into the genetic structure of European-Americans in the United States.
0279T
The timing and history of Neandertal gene flow into modern humans.
S. Sankararaman1,2, N. Patterson1, S. Pálrő2, D. Reich1,2,3
1) Harvard Medical School, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
Previous analyses of modern human variation in conjunction with the Neandertal genome have revealed that Neandertals contributed 1-4% of the genes of non-Africans with the time of last gene flow to ~86,000 years before present. Nevertheless, many aspects of the joint demographic history of modern humans and Neandertals are unclear. We present multiple analyses that reveal details of the early history of modern humans since their dispersal out of Africa.
1. We analyze the difference between two allele frequency spectra in non-Africans: the spectrum conditioned on Neandertals carrying a derived allele while Denisovans carry the ancestral allele and the spectrum conditioned on Denisovans carrying a derived allele while Neandertals carry the ancestral allele. This difference spectrum allows us to study the drift since Neandertal gene flow under a simple model of neutral evolution in a panmictic population even when other details of the history before gene flow are unknown. Applying this procedure to the genotypes called in the 1000 Genomes Project data, we estimate the drift since admixture in Europeans of about 0.065 and about 0.105 in East Asians. These estimates are quite close to those in the European and East Asian populations since they diverged, implying that the Neandertal gene flow occurred close to the time of split of the ancestral populations.
2. Assuming only one Neandertal gene flow event in the common ancestry of Europeans and East Asians, we estimate the drift since gene flow in the common ancestral population. We show that an upper bound on this shared drift of 0.010 is not consistent with this model because there is a greater drift in Europe than in Asia. The Neandertal ancestry obtained from a map of Neandertal ancestry in Eurasians, to estimate the number of generations and effective population size in the period immediately after gene flow. These analyses suggest that only a few dozen Neandertals may have contributed to the majority of Neandertal ancestry in non-Africans today.
3. We use the genetic drift shared between Europeans and East Asians, in conjunction with the observation of large regions deficient in Neandertal ancestry obtained from a map of Neandertal ancestry in Eurasians, to estimate the number of generations and effective population size in the period immediately after gene flow. These analyses suggest that only a few dozen Neandertals may have contributed to the majority of Neandertal ancestry in non-Africans today.

0281W
Rare Variant Stratification in small geographic areas. S. Zollner1,2, M. Mueller-Nurasyid1,2,3, M. Zawistowski4,5, R. Reppell6, J. Novembre2, K. Wobeser7, A. Peters7, H. Graff2, K. Straubel4,5,1) Dept Biostatistics, Univ Michigan, Ann Arbor, MI; 2) Dept Psychiatry, Univ Michigan, Ann Arbor, MI; 3) Dept of Medicine I, LMU, Munich; 4) Chair of Genetic Epidemiology, LMU, Munich; 5) Institute of Genetic Epidemiology, Helmholtz Zentrum München Germany; 6) Dept Human Genetics, Univ Chicago, IL; 7) Institute of Epidemiology II, Helmholtz Zentrum München Germany; 8) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München Germany.
The geographic distribution of rare variants provides important insights into human population history and evolutionary forces affecting our genome. Moreover, differences in rare variant counts may affect population-based association studies of rare variants; accumulation tests that jointly analyze multiple rare variants are vulnerable to confounding by very subtle population stratification. Population genetic theory suggests that rare variants are affected by small differences in migratory patterns, as such the differences between urban and rural human populations. However, little data is available to assess the degree of population differentiation in small geographic areas. To explore this question, we analyze samples from the KORA study collecting ~2400 individuals of German ancestry from the German city of Augsburg and 16 surrounding sampling locations. The sampling locations are towns and villages with census sizes between 3,000 and 30,000 located in a geographic region of ~1,200 sqkm. We evaluate the genetic differentiation between sampling locations using allele sharing, the normalized probability that two individuals sharing an allele come from a different sampling location. We show that for common and uncommon variants (minor allele frequency >0.005), allele sharing in the study area is near 1, indicating that genotype frequencies do not differ for rare variants in Europe. For rare variants, some of the same sampling locations differ from all other sampling locations and the level of differentiation does not depend on the geographic distance between sampling locations. Instead, the average allele sharing is clearly correlated with the census size of a sampling location; locations with small census size are more genetically differentiated than locations with large census size. The degree of population differentiation observed in this sample is predicted to be sufficient to generate inflated Type I error rate in accumulation tests. Tests that model causal and protective variants are especially susceptible to this type of population differentiation. Moreover, we explore the ability of typical correction strategies such as PCA and case-control matching to correct for the observed stratification.

0280F
A method for computing the exact distribution of the genealogical history of a sample derived from a structured population. M. Uyenoyama, S. Kumagai.
Biology, Duke Univ, Durham, NC.
Likelihood-based inference methods designed to analyze variation in genes sampled from structured populations require a means of accommodating changes in rates of evolutionary processes, particularly coalescence and migration. Population structure increases the dimensions of the genealogy of a sample, and migration modifies the subsets of lineages among which coalescence can occur. Some inference methods require detailed simulation of migration events to accommodate such changes. Here, we present a simple method for determining the probability of a gene genealogy under a specified model. This probability requires specification of the probability of all possible states of ancestral lineages corresponding to the nodes in a gene genealogy and the density of the time interval between nodes. We describe the qualitative features of the distribution of gene genealogies, including factors that influence the geographical location of the most recent common ancestor and the nature of the distribution of internode lengths.

0282T
Response of carrier frequency to a population bottleneck can quantify the amount of recessive variation. D.J. Balick1,2, R. Do3, D.E. Reich4,5,6, S.R. Sunyaev7,8,1) Division of Genetics, Brigham & Women’s Hospital, Harvard Medical School, Boston, Massachusetts, 02115; 2) Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142; 3) 2 Center for Human Genetics Research, Massachusetts General Hospital, Boston, Massachusetts, 02114; 4) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115.
Given the rich demographic history of humans, understanding the population dynamics in light of these differences has often been a confounding factor in identifying human evolutionary parameters, such as the strength and mode of selection. Here we present a rare events test for recessive mutations per individual subjected to a founder event to the same statistic as the previously proposed selective sweeps test, which we can exploit disparate population histories to extract a detailed understanding of crucial evolutionary parameters. In the presence of a recent bottleneck and subsequent re-expansion, the founder’s effect responds differently to distinct modes of selection, for example equilibrating more quickly from an increased recessive burden than a comparable recessive burden. The genetic variation in the pre-bottleneck equilibrated population, as measured by the average number of deleterious alleles per individual, is exceeded by the bottlenecked population when all deleterious mutations act additively, but exhibits a relative depletion under recessive selection. As a result, one can extract the dominant form of selection by comparing an easily measurable feature of a population. We introduce a methodology for probing the fraction of polymorphic alleles in a population that act recessively. We analyze the transient dynamics of the site frequency spectrum of a population in response to a bottleneck, and find a qualitative distinction between additive and recessively acting alleles. By comparing the average number of polymorphic mutations per individual subjected to a founder event to the same statistic in an equilibrium population, we are able to determine the fraction of the variation that acts recessively. The transient dynamics in response to a bottleneck are explored in detail, and a crossover is identified in the recessive fraction as a function of the dominance coefficient and the strength of selection. Under optimal epistasis, we determine bounds on this ratio. This provides a simple statistical test for the presence of recessive variation on the whole genome level, or among a collection of functionally or medically relevant loci, and can be used to parse the mode of selection among existing regional variation on the local level.
Whole genome sequencing informs genealogy and the search for genetic modifiers in a PSEN1 E280A early-onset familial Alzheimer’s disease cohort. H.C. Cox1, M. Lalli2, 5, O.E. Maukinen1, H. Li1, M. Brunckow1, M.L. Arcilia1, 2, 4, G. Garcia1, L. Madrigal6, S. Moreno5, K.S. Kosik2, 3, 4, L. Hood1, J. Roach1, G. Glusman1, F. Lopera5, 1) Institute for Systems Biology, Seattle WA 98109, USA; 2) Neuroscience Research Institute, University of California at Santa Barbara, CA, USA; 3) Program in Biomedical Science and Engineering, University of California at Santa Barbara, Santa Barbara, California; 4) University of California at Santa Barbara, Department of Molecular, Cellular, and Developmental Biology, Santa Barbara, California; 5) Grupo de Neurociencias de Antioquia, Universidad de Antioquia, Medellin, Colombia.

Mutations in the presenilin-1 gene (PSEN1) can promote progressive cerebral deposition of APP, resulting in Alzheimer’s disease (AD). In particular, carriers of the ‘Pasia mutation’, a fully penetrant and dominantly inherited amino acid change (p.Glu280Ala) in PSEN1, suffer early-onset familial AD. This mutation is highly prevalent in Antioquia, Colombia, where over 1000 carriers have been identified. The mean age of onset of mild cognitive decline and dementia in carriers is 44 and 49 years, respectively. The variation in these traits implicates additional genetic factors influencing disease progression. This population provides a unique opportunity to identify genetic modifiers underlying AD. The origin of this mutation is believed to be a single early 17th century male founder during Spanish colonization of Colombia. A comprehensive genealogy with over 3000 descendents of this male founder has been previously described. The genealogy is characterized by geographic isolation, consanguinity and admixture of native Amerindian woman with Spanish Conquistadors and individuals of native African ancestry due to the Atlantic slave trade. We have obtained whole-genome sequences for 102 individuals [from this isolated population, including 74 carriers]. We are analyzing these using custom workflows and the Ingenuity Variant Analysis platform (www.ingenuity.com/variants). We computed pairwise relationships among all sequenced individuals. We determined mitochondrial haplotypes for all individuals and Y-chromosome haplotypes for all males. We observed mitochondrial haplogroups A, B, C, L and T, which are frequent in South and Central America, Africa and Eastern Europe. We used these lines of evidence to reconstruct the extended genealogy and characterize ancestry. This study provides a unique opportunity to investigate the etiology of AD in one of the largest known AD kindreds. The reconstruction of the founder genealogy provides a comprehensive and powerful resource for whole genome sequence analysis that will inform the search for modifiers of early-onset AD. [Funding: Université du Luxembourg, www.en.uni.lu/lcsb].

Integration of low coverage whole genome sequence from the Kuusamo Finnish isolate with 1000 Genomes Project data provides an improved population reference panel. AP. Sarin1, 3, 4, K. Palin1, 2, I. Zara4, K. Rehnstrom1, 3, M. Perola1, 2, 5, V. Salomaa6, A. Palotie1, 2, 7, S. Ripatti1, 2, 3, 4, 5, R. Durbin1, 5, 7, 8, 7) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland; 3) Genome-Scale Biology Program, University of Helsinki, 00140 Helsinki, Finland; 4) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 5) Estonian Genome Center, University of Tartu, Estonia; 6) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Finland; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 8) Hjelt Institute, University of Helsinki, Helsinki, Finland.

Understanding the genetic variation in population isolates is a valuable resource for genetic studies aiming to map rare and common variation associated to complex traits and common diseases. The reduced genetic complexity in such populations leads to better genotype imputation accuracy when using a population-specific reference panel and the presence of variants rare in the general population and pushed to higher frequency by genetic drift in the isolate offers the opportunity to better represent those variants in reference panels. With the aim of evaluating the potential for imputation of including samples from an isolated sub-population in a more general population reference panel, we sequenced the whole genome of 400 healthy individuals and the whole exome of 24 additional individuals from Kuusamo, a population sub-isolate founded roughly 350 years ago in northeastern Finland. The Kuusamo population was settled by 34 families in the 1680s, and reached the present-day populations size of ~20,000 individuals without significant immigration until recently. Our simulations suggested that 400 individuals cover over 97% of all chromosomes in Kuusamo. Whole genome sequencing (WGS) was done at 4.8x on average with 100bp paired end protocol with the Illumina platform, and whole exome sequencing was done at ~37x coverage on average with the Agilent Sure Select Enrichment Kit at the Wellcome Trust Sanger Institute. Of 9,719 coding variants found in the median exome sequence, 9,595 (98.5%) are present in 1000 Genomes phase 1 (1000GP), and 98 (1.0%) more in the 400 WGS calls set. Fifteen between Kuusamo and Finnish and between Kuusamo and Europeans from 1000GP are 0.006 and 0.013 respectively. We also evaluated the proportion of IBD segments shared across individuals and the percentage of individuals with at least one surrogate parent, as described in Kong at. al, Nature Genetics 2008, using the Systematic Long Range Phasing (SLRP) software (Palin et al, Genetic Epidemiology 2011). We evaluated imputation quality of 41 additional Kuusamo samples and 40 Helsinki samples using IMPUTE2. In general imputation is better into the Kuusamo samples than the Helsinki samples. Moreover, in Kuusamo samples a combined reference panel with both Kuusamo and 1000GP haplotypes provides the best performance. Our study demonstrates the value of combining outbred, isolated and sub-isolated population data to increase the amount of variability captured through imputation.
2085T

Genealogical evidence of allele frequency shuffling explains regional genetic structure in a founder population. C. Bherer1, L. Excoffier1, 2, 3, M.H. Roy-Zagnon4, J. Hussin5, H. Vézina5, D. Labuda1, 6, 7, 1) Centre de recherche du Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Québec, Canada; 2) CMPG, Institute of Ecology and Evolution, University of Bern, 6, Baltzerstrasse, CH-3012 Bern, Switzerland; 3) Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; 4) Department of Epidemiology and Community Medicine, University of Ottawa, Ontario, Canada; 5) BALSAC Project, Université du Québec à Chicoutimi, Québec, Canada; 6) Department of Pediatrics, Université de Montréal, Québec, Canada.

Founding events, allele frequency changes and regional demographic histories are responsible for the increase in frequency of rare disease-causing variants, and if this demography has affected the site frequency spectrum (SFS). By simulating the Mendelian transmission of alleles in the Quebec population. We studied allele frequency changes following the initial founding settlement and its subsequent regional spatial expansion. We used ascending genealogies of 2,221 individuals representing eight Quebec regions. The total genealogy includes 153,447 ancestors linked to 8,834 founders. Allele dropping (AD) simulations, starting from an equilibrium SFS among founders, led to significant loss for alleles below 1%, but no changes for the larger frequency classes. Using Tajima’s D equilibrium test, a significant deficit of rare alleles was found in the whole Quebec sample, in the North-Eastern and Eastern regions, whereas an excess of rare alleles was observed in the other regions. Simulations suggest that regional demographic histories had opposite effects. FST analysis showed significant genetic structuring of the regional populations. We further quantified frequency changes for different initial allele frequencies among founders. For both rare and common alleles, the increase in frequency changes was significant between regions, demonstrating differential shuffling effect of allele frequencies among regional populations. We observed extensive variance in genetic contribution among founders. In the North-Western, North-Eastern and Eastern regions, we find that a unique allele carried by a single founder could have increased in frequency up to 5% until now, thus potentially explaining the clinical founder effect of specific Mendelian disorders observed in the North-East. Our results demonstrate that regional demographic history resulted in regional founder effects and genetic structure over a very short evolutionary time.

2086F

Genetic characterisation of two Greek population isolates. K. Hatziko-Goridis, A. Farmaki, N.W. Raymer, C. Ch. Kiagiadakis, E. Tsafaktakis, N. Ntalla6, N. Karaleftheri, G. Dedoussis, E. Zeggini,1 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Department of Genomics, Karolinska University Hospitals, Stockholm, Sweden; 3) Centre de recherche du Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Québec, Canada; 4) Department of Genetics, University of California, San Diego, La Jolla, CA, USA; 5) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 6) Department of Anatomy, University of Athens, Athens, Greece; 7) Pathology, Montefiore Medical Center, Bronx, NY.

By collecting >2,000 samples from two isolated populations in Greece: the Kentavros village in the Attica district, and the Pomak isolate in the North of Greece, and perform haplotype analysis using the PHASE software. Results: Among our 1,494 AJ samples, 52 were shared by the Kentavros village and the Pomak isolate. Genetic association studies of low-frequency and rare variants can be empowered by focusing on isolated populations. It is important to genetically characterize population isolates for substructure and recent admixture events as these may give rise to spurious associations. Under the auspices of the HELenic Isolated Cohorts study (HELIC; www.helic.org) we have collected >3,000 samples from two isolated populations in Greece: the Pomak villages (HELIC Pomak), a set of religiously-isolated mountainous villages in the North of Greece, and a group of isolated villages known as MANOLIS. All samples have information on anthropometric, cardiometabolic, biochemical, haematological and diet-related traits. 1,500 individuals from each population isolate have been typed on the Illumina OmniExpress and Human Exome Beadchip platforms. Multidimensional scaling analysis with the 1000 Genomes Project data shows similarities of the two population isolates with Mediterranean populations such as the Tuscan from Italy and the Iberians from Spain. We also observe evidence for structure within the isolates, with the Kentavros village in the Podxorras group, and the Pomak group, which is located in the MANOLIS isolate. The authors have implemented a method to determine the degree of isolationized in these populations we estimated the proportion of individuals with at least one ‘surrogate parent’ (using only the subset of samples with pairwise r-hat<0.2) and compared this to an outbred Greek population from the TEENAGE study. We find that 707 unrelated adolescents from the Attica district. We find that for random regions in the genome the proportion of individuals with at least one surrogate parent in the MANOLIS isolate is >60% and in the Pomak isolate is >65% compared to ~1% in the outbred Greek population. These results establish two population isolates as isolates and provide some insights into the genomic architecture of Greek populations, which have not been previously characterized.

2087W

Risks for Mendelian Disorders in the Bronx. K. Upadhya1, H. Oster1, 2, C. Oddoux2. 1) Pathology, Albert Einstein College of Medicine, Bronx, NY; 2) Pathology, Montefiore Medical Center, Bronx, NY.

The Bronx, a borough of 1.4 million people, is the third most genetically diverse county in the United States. Of the current residents, 51% are Hispanic-Latino (of whom Puerto Ricans and Dominicans are the largest ethnic groups) and 35% are African-American. To understand the risks of Mendelian disorders among these populations, Affymetrix Axiom Exome 319K arrays were analyzed for 192 members of each group, followed by variant calling using Affymetrix Genotyping Console software. Following quality control, calls were made for 305,519 probes. The called probes were mapped for annotated variants by cross-referencing to Online Mendelian Inheritance in Man, resulting in 1,440 known alleles, of which 818 were polymorphic and 338 were pathogenic for Mendelian disorders. These alleles were also cross referenced against 14 other World populations, including an independent set of Puerto Ricans (demonstrated to overlap with Bronx Puerto Ricans by PCA analysis). Among these alleles, 157 were shared among all 3 populations, 32 were shared by Puerto Ricans and Dominicans, 22 were shared by Puerto Ricans and African Americans and 12 were shared by Dominicans and African Americans. In singleton populations, 37 were found only in Dominicans and 30 were found only African Americans. Among the conditions previously described in these populations were prostate cancer susceptibility and progression (EPHB2 p.A279S), hemochromatosis (HFE p.C282Y), and hereditary amyloidosis (TTR p.I112V). Among the conditions with allele frequencies ≥1% in at least one population were cystic fibrosis (CFTR p.F508del), spastic paraplegia 44 (GJC2 p.196M), susceptibility to diffuse gastric cancer (CDH1 p.A617T), Usher syndrome 3C (USH1C p.V130I), cone-rod dystrophy 3 (ABCA4 p.L1201R), fructose intolerance (ALDOB p.A150P), Fuch’s endothelial 6 corneal dystrophy (ZEB1 p.G841S) and susceptibility to intestinal carcinoid tumors (SDHD p.H50R). Thus, these studies provide insights into previously unrecognized mutations for Mendelian disorders among the populations of the Bronx.

Whole-exome or gene targeted resequencing in hundreds to thousands of individuals has shown that the majority of genetic variants are at low frequency in human populations. Rare variants are enriched for functional mutations and are expected to explain an important fraction of the genetic etiology of human disease, therefore having a potential medical interest. In this work, we analyze the whole-exome sequences of French-Canadian individuals, a founder population with a unique demographic history that includes an original population bottleneck less than 20 generations ago, followed by a demographic explosion, and the whole exomes of French individuals sampled from France. We show that in less than 20 generations of genetic isolation from the French population, the genetic pool of French-Canadians shows reduced levels of diversity, higher homozygosity, and an excess of rare variants with low variant sharing with Europeans. Furthermore, the French Canadian population has accumulated a large number of putative damaging functional variants, that may impact upon the fitness of the population and could partially explain the increased incidence of genetic disease in the province. Our results highlight the importance of population demography on genetic fitness and the contribution of rare variants to the human genetic variation landscape, emphasizing the need for deep cataloguing of genetic variants by resequencing worldwide human populations in order to truly assess disease risk.

Whole exome sequencing in families with rare perinatal onset immunodysregulatory diseases present with fever and systemic inflammation. Z. Deng, Y. Liu, A. Almeida de Jesus, H. Sun, M. Gadina, R. Goldbach-Mansky. Intramural Research Program, NIAMS, National Institute of Health, Bethesda, MD.

WES (Whole Exome Sequencing) has increasingly become the tool of choice in translational research, providing molecular diagnosis in Mendelian diseases and identifying important genes in key biological pathways. Here we report using WES to investigate the molecular basis of a group of rare immunodysregulatory diseases that are characterized by perinatal onset fever and systemic inflammation. We have developed a bioinformatics pipeline to process WES data and an integrated workflow to analyze variants in family trios or quartets. Using the pipeline, we were able to assess the quality of WES data and identify discrepancies in sample gender and family relatedness. The number of coding variants per sample in our WES studies ranged from 19,000 to 25,000, correlating with exome coverage and sample relatedness. The number of coding variants per sample in our WES studies ranged from 19,000 to 25,000, correlating with exome coverage and sample relatedness. We show that in less than 20 generations of genetic isolation from the French population, the genetic pool of French-Canadians shows reduced levels of diversity, higher homozygosity, and an excess of rare variants with low variant sharing with Europeans. Furthermore, the French Canadian population has accumulated a large number of putative damaging functional variants, that may impact upon the fitness of the population and could partially explain the increased incidence of genetic disease in the province. Our results highlight the importance of population demography on genetic fitness and the contribution of rare variants to the human genetic variation landscape, emphasizing the need for deep cataloguing of genetic variants by resequencing worldwide human populations in order to truly assess disease risk.

The impact of population demographics and selection on the genetic architecture of complex traits. K.E. Lohmueller. Ecology and Evolutionary Biology, University of California, Los Angeles, CA.

The medical genetics community is moving into an era of using exome sequencing to identify low-frequency (<1%) coding region variants associated with complex traits. At the design stage of such experiments, it is critical to understand how many and which individuals to sequence. One important factor that has not been explored fully in previous studies is the demographic history of the sequenced population. Recent studies of thousands of individuals have found evidence of extreme recent population growth in several populations. Here I investigate the effect of this growth on the power to identify low-frequency variants associated with complex disease. In particular, I perform forward simulations using realistic demographic models for different human populations that take into account various levels of recent population growth. The simulations also include a distribution of selective effects for deleterious mutations. Phenotypes are simulated using a liability threshold model where a mutation’s effect on the phenotype is proportional to its effect on fitness. First, my simulations show that recent population growth leads to a proportional increase in deleterious amino acid changing polymorphisms, especially if the growth occurred so recently that purifying selection has not had sufficient time to remove the deleterious mutations. Second, if a mutation’s effect on fitness is correlated with its effect on the trait, then rare variants account for more of the additive genetic variance of the trait in a recently expanded population than in a population that did not expand. Third, in line with the previous result, I show that recent population growth can increase the genetic heterogeneity of disease. Specifically, a sample of diseased individuals from a population that has undergone recent growth is expected to contain more distinct causal mutations than a sample of the same number of cases from a population that has not expanded. If the causal mutations are scattered across many distinct genes, this increased heterogeneity due to population growth will reduce the power of commonly used gene-based rare variant association tests. These findings suggest that careful consideration of recent population history is essential for designing optimal medical sequencing studies.

Regional rare allele sharing in a homogeneous, geographically clustered population. S.L. Pultit, A. Menelau, C.C. Elbers, L.C. Francioni, P.I.W. de Bakker, The Genome of the Netherlands Consortium. 1) University Medical Center Utrecht, Utrecht, The Netherlands; 2) Medical and Population Genetics Program, Broad Institute, Boston, MA, USA.

Next-generation sequencing has yielded unprecedented cataloguing of variants across the frequency spectrum. Although population structure in common variation (CV) is well studied, the exploration of population structure due to rare variants (RVs) has only just begun. Some studies suggest that RV structure is distinct from that of CV while others have found that it can be captured by principal components (PCs) calculated using common SNPs. A broader understanding of population structure is needed as we increasingly focus on the role of RVs in complex traits. To elucidate the architecture of RVs, we used Genome of the Netherlands (GoNL) data of 250 Dutch trios sequenced at 14x coverage. We specifically analyzed f2 variants, mutations appearing exactly twice in GoNL (N=330 unrelateds evenly sampled from 11 provinces). Since f2 are rare and likely to be newer mutations, they can illuminate demographic history. We annotated f2 allele carriers and their province of birth and found that within-province sharing was strongest; on average, both f2 alleles were 1.46x more likely to be found in the same province. After dividing the Netherlands into northern, central, and southern regions, we tested if within-region sharing of f2 was stronger in the north or south, areas with historically less migration, compared to the central provinces, areas with more migratory populations, the country’s largest city (Amsterdam), and biggest transport (Rotterdam). We observed strong sharing in the northern vs. central regions (p<10^-200), but no excess sharing for the southern vs. central regions (p=0.11). These observations are consistent with Dutch history: for centuries, flooding in the northern regions prevented migration into or out of the north and populations primarily settled in the north. Conversely, central regions also experienced higher CV and RV structure overlap by projecting the samples onto CV PCs. While PC1 captures a North-South cline, the northern samples (with much stronger f2 sharing) are no more tightly clustered in PC space (mean distance=0.072) than the central samples (mean dist=0.065, p=0.99). Our results suggest that RV structure, even in a homogeneous population from a small geographic area (41,000 km2), is not fully captured by PCs, which has important implications for RV disease studies such as those employing the exome chip.
Estimation of Migration Rates and Patterns Based on Distributions of Rare Variants. R. Rothwell, M. Ehm, P. St. Jean, M. Nelson, J. Novembre, S. Zöllner. 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Quantitative Sciences, GlaxoSmithKline, RTP, NC; 3) Human Genetics, University of Chicago, Chicago, IL.

Careful modeling of the relationships between populations is essential to our understanding of population diversity. Most current models, however, ignore historical changes in migration rate, modeling gene flow as a constant parameter. The growing availability of sequencing data allows estimating temporal changes in these relationships. Identifying such changes may quantify the impact of historical events on human diversity. In this study, we develop a method for using rare variants to identify changing migration patterns. The distribution of rare alleles among multiple populations depends on the migration rate between those populations in the historic time since the mutation event that generated the variant. Rare variants that arose recently are only affected by recent migration events and thus allow an estimate of the migration rate in the recent past. Common variants that arose further in the past allow an estimate of the average migration rate over the time since they arose. Using this intuition, we develop a likelihood function based on the distribution of alleles, where we obtain one estimate of the migration rate each from all variants with a given minor allele count. By comparing different estimates of migration across variants with different minor allele count, we generate a temporal outline of migration rate over time. We evaluate the performance of our method on simulated data of populations with changing migration patterns. For changes as recent as 20 generations in the past, our algorithm identifies a clear temporal pattern. We also show that this methodology can be adjusted to model population characteristics including changes in effective population size and is robust to minor model misspecifications. We apply this methodology to sequence data from the exons and flanking regions of 202 target genes (2220 exons total), generated from European and African American samples. In the African American population, we observe sharply increased migration in recent years from European populations, reflecting a pattern consistent with the transatlantic slave trade. In the European populations, we find generally high gene flow with slightly lower migration rates from the more geographically distant populations. In the Nordic-Baltic region, we observe temporal trends indicating widespread Scandinavian expansion across Europe and subsequent reduction in migration. In Central and Western Europe, we find increasing European movement in recent years.

Exome sequencing revealing unique genetic profile of Quebec Nunavik Inuit population. S. Zhou, L. Xiong, A. Ambalavanan, A. Dionne-Laporte, O. Spiegelman, E. Henning, O. Diao, C. Bourassa, N. Dupré, M.-P. Dubé, P. Dion, G. Rouleau. 1) Centre of Excellence in Neuroscience of Université de Montréal (CENUM), CHUM Research Center and the Department of Medicine, Montréal; 2) Centre de recherche Fernand-Seguin, Hôpital Louis H. Lafontaine, Montréal; 3) Department of Human Genetics, McGill University; 4) Department of Neurology, Université Laval, Québec; 5) Pharmacogenomics Centre, Montreal Heart Institute, Université de Montréal, Montréal; 6) Montreal Neurological Institute and Hospital, McGill University, Montréal.

Background: Nunavik comprises the northern part of Quebec, with 90% of its inhabitants being Inuit. The modern Inuit of Nunavik came from the Thule people which originated from coastal Alaska around 1000 AD and traveled eastward along the Arctic tree line, toward the Northwest Territories (NWT), Nunavut, Greenland, and reached Nunavik around 1500 AD as they became the ancestors of current Inuit residents. In our study, we chose to use a high throughput parallel sequencing approach to generate the specific genetic profile of the Inuit population for the first time. Materials and Methods: We performed exome sequencing on 101 Nunavik Inuit from 11 out of 15 villages in Nunavik, using Agilent SureSelect V4 capture kit and Illumina HiSeq platform. Standardized data processing is used to extract all exonic variants. Results: We performed preliminary analysis in order to identify Inuit specific novel or rare nonsynonymous variants. The analysis yielded 62 protein changing variants which have allele frequency over 0.5 in Inuit while less than 0.01 in other populations. Among these, a variant p.P479L in the Carnitine Palmitoyltransferase 1A (CPT1A) gene is the most prominent one. 88 individuals from our Inuit cohort carry the homozygote, and with the lack of the homozygote wild-type, the p.P479L allele frequency in Nunavik reaches 0.94. This frequency appears to be the highest when compared to other Inuit populations: 0.43 in Alaska native newborns, 0.44 in NWT Inuit, 0.77 in Nunavut and 0.73 in Greenland Inuit. Interestingly, the increasing frequency of this variant from Alaska to Greenland and the northeast of Canada seems to correlate with the migration pattern of Inuit over the course of a thousand years. Discussion: CPT1A is a key regulator of fatty acid metabolism, and the p.P479L allele is known to be only present in Inuit populations. This variant has been reported to be associated with impaired fasting tolerance in Alaska, and increased risk of sudden infant death (SIDS) in Nunavut. However, in Greenland Inuit the p.P479L allele is linked with higher levels of HDL-cholesterol and apoA-I, which are considered to be a protective allele against cardiovascular risk and infection. The interesting diverse phenotypic variation associated with this allele and its high frequency among older Nunavik adults (mean age 52) suggest the same variant may contribute to different physiological activities in Inuit during different life stages and from different environments.
Application of standard principal components analysis (PCA) methods to sequence data does not perform well because of (1) pervasive linkage disequilibrium (LD) and (2) large numbers of rare variants. When applied to 0.78 Mb of targeted sequence data from 518 individuals of European ancestry, we examined the performance of ancestry inference, we considered exclusion of rare variants rather than LD regression (regressing out surrounding SNPs from the performance of ancestry inference, we used LD to handle rare variants. To correct a distribution of the variance in true population labels (r^2 = 0.023). To improve the performance of ancestry inference, we examined the effectiveness of several methods to correct for LD and handle rare variants. To correct for LD, we considered LD pruning, LD shrinkage (reweighting by LD to surrounding SNPs) or LD regression (regressing out surrounding SNPs from each SNP). All of these methods improved results, with LD regression performing best. To handle rare variants, we considered exclusion of rare variants, exome capture, or LD correction. Application of LD correction and inclusion of rare variants resulted in far greater accuracy (r^2 = 0.99), sufficient to correct for population stratification in association analysis. Overall, our results suggest that effective ancestry inference from targeted sequence data or targeted high-density genotyping data requires appropriate treatment of LD and rare variants. In particular, rare variants can improve ancestry inference, and should not be universally removed.
2098F
A low frequency of copy number polymorphism for the NPEPPS, POLR2J4, and PCDH13 genes in a sample of admixed Brazilian population. D. Jimenez1, T. Lins2, P. Taveira3, R. Pereira1,2. 1) Universidade Católica de Brasilia, Brasília, DF, Brazil; 2) Patologia Molecular, Universidade de Brasilia, Brasília, DF, Brasil.

Copy number variants (CNV) represent an important source of variation in the human genome, with some embedded genes differently distributed among human population groups. Therefore, it is necessary to understand the distribution of CNV within and between populations, especially those with admixed ancestry, such as the Brazilians. The aim of the study was to investigate the variability of a set of CNV-embedded genes in a sample of the Brazilian population. The CNV-embedded genes were chosen based on previous published data showing that they have differential copy variation distribution between population samples from Africa and Europe. Three genes (NPEPPS, POLR2J4, and PCDH13) were investigated by real-time qPCR using TaqMan® assays (Applied Biosystems, Foster city, CA) in a sample of 96 Brazilian individuals previously classified by genetic ancestry informative markers. A low variability in the selected genes was identified, showing 2 copies in frequency of 0.927, 0.990 and 0.979, respectively. Mean distribution ± standard deviation of ancestry according to European, African and Native American proportion was, respectively, 0.702 ± 0.232, 0.189 ± 0.158 and 0.109 ± 0.147, for overall sample. Due to low variability, the genomic ancestry was correlated to copy number (diploid = 2 copies versus non-diploid) only in NPEPPS gene, but no correlation was observed (ρ = 0.257, 0.188 and 0.498, respectively for ancestry estimates). In conclusion, the results provided an overview of the corresponding frequency of gene copy number variation in a sample of the Brazilian population, serving as reference for further population studies, which maybe correlate these polymorphisms with other phenotypic differences. Nevertheless there is a need for upcoming research to revise these data by using more assays nearby the genes to search for specific breakpoints or use other specific methods, such as for the hypotheses and second generation sequencing, to confirm and validate the copy number polymorphisms.

2099W
Amylase gene copy number polymorphism in ethnically admixed sample from Brazil. T. Lins1, D. Jimenez2, P. Taveira2, R. Pereira1,2. 1) Patologia Molecular, Universidade de Brasilia, Brasilia, DF, Brazil; 2) Ciencias Genomicas e Biotecnologia, Universidade Catolica de Brasilia, Brasilia, DF, Brazil.

Salivary amylase gene (AMY1) copy number variation (CNV) represent a significant variation among individuals and evolved independently in different human population groups worldwide. The AMY1 genetic variation seems to be related with eating habits and dietary consumption of starch as the AMY1 gene copy number, the concentration and activity of salivary amylase enzyme are positively correlated to the starch content diet (high or low starch diet). Therefore, it is necessary to understand the distribution of CNV within and between populations, especially those with admixed ancestry, such as the Brazilians. The aim of the study was to investigate the variability of the AMY1 gene in a sample of 96 Brazilian individuals previously classified by genetic ancestry informative markers. The polymorphism was investigated by real-time qPCR using Taqman® assay (Applied Biosystems, Foster city, CA). The copy number was evaluated with a range of 1 to 8 copies. The major copy number was identified for 2 copies (47.9%), followed by 3 copies (20.8%) and 4 copies (18.8%). Mean distribution ± standard deviation of ancestry according to European, African and Native American proportion was, respectively, 0.702 ± 0.232, 0.189 ± 0.158 and 0.109 ± 0.147, for overall sample. A T-test was performed to compare estimated genomic ancestry proportion among 2 copies and gain of copies (≥3 copies), but no statistical correlation was observed (ρ = 0.449, 0.494, 0.412, respectively). Despite the low copy number compared to other worldwide populations, the correlation of the polymorphisms with other phenotypic differences must be investigated, such as salivary amylase enzyme concentration and activity, as well as the nutritional status and starch content diet. Since the AMY1 is highly variable, an increased number of individuals in the Brazilian population should be further evaluated to screen a wider range of variation. The results provided an overview of the corresponding frequency of AMY1 gene copy number variation in a sample of the Brazilian population serving as reference for further population and nutrigenomic studies.

2100T
Mitochondrial Genome Database for Saudi Community, I. Alabdulkareem1, M. Albalwi1, A. Alharbi2, A. Alghamdi3, B. Alhamad4, M. Aljumah4. 1) KAIMRC, National Guard Health Affairs, Riyadh, Riyadh, Saudi Arabia; 2) King Abdulaziz Medical City for National Guard, Pathology Department, Riyadh, Saudi Arabia.

This study is aiming to create an electronic database for all Non Mendelian inheritance among Saudi populations that will be available for all researchers and health providers at the country. Phase one of this investigations is focusing on the establishment of the mitochondria genome sequencing in healthy (mtDNA), mtDNA as extracellular components that do not subdue to Mendelian inheritance mode that is genetically considered as complete different in terms of evolutionary origin since many of the evaluation scientists thought that mtDNA may be derivative from the circular genomes of the bacteria that were engulfed by the early ancestors (endosymbiotic hypothesis). mtDNA was sequenced using the solid 5500™ and validate by Ion Personal Genome Machine® [PGM™] System. Studies showed that the Arab assembled mitochondrial genome contains 16,570 bp as compared to GenBank reference genome with 16,571 bp. Further analysis in mitochondrial genome detected 46 differences, with almost single point mutations and only a deletion of length one. Homology studies on 3,132 known mitochondrial sequences from NCBI with sequence length between 16,000 and 17,000 bp revealed more than 90% matches sequence to accession number EF184636 sequence. Reported single nucleotide polymorphisms (SNPs) are rs3927813 rs3928312 rs1599988 rs3021088 rs3929989 rs1978028 rs121434446 rs1981459 rs3902407 rs3915952 rs3926883 rs3899498 rs2835780 rs3135031 rs34799580. More subjects from healthy and diseased will be included in this investigation that will facilitate more understanding of many unexplained chronic diseases among the populations.

2101F
Evolutionary insights into genetic and environmental factors for Crohn's disease in Ryuku and Northern Kyushu. S. Nakagome1, H. Chirine2, H. Saito3, W. Suda3, A. Itahara4, A. Hokama4, F. Kinjo5, J. Fujita5, Y. Takeyama5, S. Sakisaka5, T. Matsu6, K. Kidd7, K. Kida7, S. Kawamura7, T. Hanhara8, R. Kimura9, H. Ishida10, M. Morita10, M. Hattori11, S. Mano11, H. Oota11. 1) The Institute of Statistical Mathematics, Tokyo, Japan; 2) University Hospital, Faculty of Medicine, University of the Ryukyus, Okinawa, Japan; 3) Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan; 4) Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan; 5) Fukukao University Faculty of Medicine, Fukukao, Japan; 6) Fukukao University Chikushi Hospital, Fukukao, Japan; 7) Yale University School of Medicine, New Haven, CT; 8) Kitasato University School of Medicine, Kanagawa, Japan; 9) Azabu University, Kanagawa, Japan.

Crohn's disease (CD) involves a chronic inflammation in the intestinal tracts, which are caused by complex interaction between multiple genetic and environmental factors. More than seventy loci have been reported as susceptibility genes for CD in Europeans. However, these susceptibilities are different between Japanese and Europeans. Here, we hypothesize that there is the population-specific susceptibility even among Japanese populations. Then, we focus on the TNFSF15 locus whose association with CD has been shown in the Honshu Japanese, and investigate the susceptibility in the Ryuky and the Northern Kyushu Japanese. The allelic association of TNFSF15 with CD was shown in both of the populations, while the genotype relative risk was different between the Ryuky (recessive) and the Northern Kyushu (dominant) Japanese. Further, we found that the frequencies of the risk alleles were significantly higher in the Ryuky Japanese than in the Northern Kyushu and the Honshu Japanese, as well as in the Korean population. These results suggest that the geographic distribution of risk alleles in the TNFSF15 locus can be explained by the demographic effects under the dual-structure model of Japanese populations, and the susceptibility to CD is determined by environmental factors.
2102W
The demographic and genetic dynamics of the human sex ratio from conception to birth. S. Orzack. Fresh Pond Research Institute, Cambridge, MA.

We describe the demographic and genetic trajectory of the human sex ratio from conception to birth by analyzing data from 1) three-day-old embryos, 2) induced abortions, 3) chorionic-villus sampling, 4) amniocentesis, and 5) fetal deaths and live births. Our data set is the most comprehensive and largest ever assembled to estimate the sex ratio at conception and the sex ratio trajectory and is the first to include all of these types of data. Our estimate of the sex ratio at conception is 0.5 (proportion male), which contradicts the common belief that the sex ratio at conception is male-biased. The sex ratio among abnormal embryos is male-biased and the sex ratio among normal embryos is female-biased. These biases are associated with the abnormal/normal state of the sex chromosomes, chromosome 15, and chromosome 17. The sex ratio of a cohort of conceptions increases for at least 10 to 15 weeks after conception. The sex ratio levels off after approximately 20 weeks and then declines slowly from 28 weeks to 35 weeks, when it declines markedly and becomes female-biased. This trajectory indicates that there is an early excess of female mortality during pregnancy followed by a later excess of male mortality, with net female mortality during pregnancy exceeding male mortality; this is a fundamental new insight into early human development.

2103T
Critical illness from pandemic influenza A/H1N1 and Streptococcus pneumonia co-infection in Nuevo León, Mexico. B. Silva1, A. Padron-Rocha1-3, D. Resendez-Perez2, M. Bermudez-de Leon2. 1) Departamento Inmunogenetica, CIBIN-IMSS, Monterrey, Nuevo León, Mexico; 2) Departamento Biologia Molecular, CIBIN-IMSS, Monterrey, Nuevo León; 3) FCB-UANL; 4) Departamento de Virologia e Immunologia, FCB-UANL.

The influenza A/H1N1 virus pandemic of 2009 started in Mexico and then spread worldwide. Bacterial infection associated with influenza is a common cause of death on seasonal and pandemic influenza. Streptococcus pneumoniae is the major bacterial pathogen associated with the influenza in the 1918 Spanish pandemic. The contribution of bacterial coinfection to critical illness associated with influenza A/H1N1 virus infection remains uncertain. The objective of this study was to determine the association of the influenza pandemic of 2009 with S pneumoniae and its contribution to critical illness. Design: Retrospective cohort study Patients: We examined nasopharyngeal swab samples (NPS) of three hundred five patients who were confirmed for all samples with the WHO approved Real Time PCR influenza A/H1N1 virus pandemic of 2009 essays. Autolysin, encoded by (lytA) gene is required for all samples with the WHO approved Real Time PCR influenza A/H1N1 virus.

The influenza A/H1N1 virus pandemic of 2009 started in Mexico and then spread worldwide. Bacterial infection associated with influenza is a common cause of death on seasonal and pandemic influenza. Streptococcus pneumoniae is the major bacterial pathogen associated with the influenza in the 1918 Spanish pandemic. The contribution of bacterial coinfection to critical illness associated with influenza A/H1N1 virus infection remains uncertain. The objective of this study was to determine the association of the influenza pandemic of 2009 with S pneumoniae and its contribution to critical illness. Design: Retrospective cohort study Patients: We examined nasopharyngeal swab samples (NPS) of three hundred five patients who were confirmed for all samples with the WHO approved Real Time PCR influenza A/H1N1 virus pandemic of 2009 essays. Autolysin, encoded by (lytA) gene is required for S pneumoniae pathogenesis and well-characterized virulence marker. Real-time PCR with sequence-specific primers and fluorescent TaqMan probes were selected to direct detection from NPS. Demographic data, comorbid conditions, illness progression and clinical outcomes were collected. Results: S pneumoniae co-infection was identified in 10.49% of patients with severe influenza A/H1N1 virus infection. Patients with co-infection were young (median 24.5 (range, 0-76) years). Co-infection was associated with increased ICU admission (75% patients). Hospital mortality was 16.23% in bacterial co-infection. The presence of S pneumoniae was correlated with severe disease. Complex viral, bacterial and host factors contribute to pathogenesis of co-infection. Reductions in morbidity and mortality are dependent on prevention with available vaccines as well as early diagnosis and treatment.

2104F
Quantitative genetic analysis of reactivity to three distinct behavioral challenges in an infant primate model of susceptibility to anxiety disorders and depression. G.L. Fawcett1, A.M. Deitmer2, D. Kay3, M. Ravendran4, J.D. Higley4, N.D. Ryan5, J.L. Cameron5,6, J. Rogers7. 1) Department of Molecular & Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 3) Department of Clinical & Health Psychology, University of Florida, Gainesville, FL; 4) Department of Psychiatry, Brigham Young University, Provo, UT; 5) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA; 6) Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR.

Prior data clearly show that susceptibility to anxiety disorders and depression is strongly influenced by genetic differences between individuals. Studies of human children in response to novel stressful environments have revealed moderate additive genetic heritability (h2=0.3) of anxious temperament, a risk factor for later adult mood disorders. Rhesus macaques (Macaca mulatta) are the most commonly used non-human primate in biomedical research, including use as models for studies of alcoholism and drug addiction, response to stress and basic aspects of neurodevelopment. Given the significance of developmental experience as well as genetics in forming temperament later in life, we conducted a behavioral genetic analysis of a large (n=428) set of pedigreed infants raised in large social groups that mimic species-typical social organization and interaction. In order to reveal individual differences in underlying temperament, we challenged infant monkeys (aged 40-158 days) with a series of standardized behavioral tests. Each of these conditions varied the environment so as to alter the mildly stressful challenge conferred to the infant: (1) novel environment without threatening stimulus with the mother sedated but present (Freeplay test), (2) novel environment without threatening stimulus without the mother present (Human Intruder Test (HIT)-Alone), and (3) novel environment with exposure to a mildly threatening stimulus (i.e., human staring at monkey) without the mother present (HIT-Stare). We observed significant heritability for willingness to move away from the sedated mother to examine a novel environment (h2=0.25±0.13, p=0.003). Infants showed heritable variation in a range of behavioral responses to such mildly stressful threats when the mother was not present (h2=0.29±0.32, p=1.5×10^{-3}-2.1×10^{-4}). By examining these different testing paradigms, we are able to identify genetic influences on reactivity to different environmental challenges. Future analyses will explore genetic and environmental interactions among infant environment and test for genetic associations between different aspects of behavioral variation and a series of specific candidate genetic pathways.

2105W

DNA methylation varies temporally across cell types in an individual, regulating gene expression crucial to biological processes. While methylation is reprogrammed during early development, it also has a heritable component and recent work has shown that this trans-generational inheritance is influenced both by genetic variation as well as environmental factors such as diet. Here we have employed a targeted padlock probe bisulfite sequencing method to characterize DNA methylation in white blood cell samples obtained from a culturally diverse set of sub-Saharan Africa individuals practicing agricultural, pastoralism, agro-pastoralism and hunting-gathering. Using a dimension reduction approach, we demonstrate that the methylation profile of the West African Pygmy population samples is distinct from other Africans even after controlling for covariates such as age and sex. We find that the methylation sites contributing to this signal are enriched near genes that play a role in growth and immunity. In addition, after adjusting methylation frequencies for variation in nearby SNPs, or methylation quantitative trait loci (methylQTL), we find that this Pygmy-specific signal still persists, suggesting an environmental component to differences in genome-wide methylation patterns of Pygmy individuals.
The myocardial infarction-associated gene PHACTR1 is controlled by MEF2, M. Beaudoin2, S. Langlois3, K. Lo4, A. NDiaye5, J.C. Tardif2,6, G. Lettre1,7,8,10,11,12,13, M. Beaudoin2, S. Langlois3, K. Lo4, A. NDiaye5, J.C. Tardif2,6, G. Lettre1,7,8,10,11,12,13, L.P. Surve14, L.Y. Sun15, R. Chagnon16, M. Gregoire17, E. X. Hu18, L. Fournier19,20,21, P. A. Hosseinib, P.A. Cleyer22, J.M. Lachin2,3, A.D. Paterson1,2,3, DCCT/EDIC Research Group. 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 3) Department of Statistical Sciences, University of Toronto, Canada; 4) Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada; 5) Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Prosserman Centre for International Health, Toronto, Ontario, Canada; 6) Division of Cardiology, Department of Medicine, University of Toronto, Toronto, Canada; 7) The Biostatistics Center, The George Washington University, Rockville, Maryland.

Elevated LDL-C is a risk factor for cardiovascular disease, a leading cause of death in T1D. Randomization to intensive therapy, aimed at achieving glycemic control in the non-diabetic range, was associated with lower LDL-C in comparison to conventionally treated subjects in the Diabetes Control and Complications Trial (DCCT). The largest genome-wide association study of lipids to date identified 37 loci associated with LDL-C using over 95,000 primarily non-diabetic subjects (Teslovich et al., 2010). Given higher glycemia in T1D, it is unknown whether these loci are similarly associated with LDL-C in T1D. Therefore, to test this hypothesis, we used longitudinal measures of LDL-C from 1304 white participants with T1D from the DCCT. Association results from linear mixed models adjusted for time, randomized treatment group and other covariates showed that the directions for the SNP effect at 35 of the 37 loci were consistent with those reported by Teslovich et al., far exceeding the null expectation of 18.5 (p<10-7). Individual variants showed evidence of association in T1D, and correcting for multiple testing only 5 (LDLR, APOE-C1-C2, SORT1, ABCG5/A, APOB) remained statistically significant (p<0.001), suggesting a larger sample is needed to confirm the other locus-specific associations. To test the hypothesis that these 37 loci jointly contribute to susceptibility of higher LDL-C in an additive fashion, we then considered the polygenic risk score method. Risk allele was defined based on the Teslovich et al sample, and risk scores were calculated for each person by multiplying the number of risk alleles by the beta estimate from Teslovich et al for each SNP and summing the product across all 37 loci. Results of this weighted polygenic risk score analysis showed a significant additive genetic effect (p<10-38). To ensure that the result was not driven by the top 5 loci in DCCT, we repeated the analysis focusing on the remaining 32 loci. The strength of the association was attenuated as expected but remained significant (p<10-15). We also tested for a SNP by TREATMENT interaction at each locus but observed only suggestive evidence for interaction at ABO (p=0.01) and ST3GAL4 (p= 0.03). In conclusion, loci ascertained in non-diabetics are similarly associated with lipids in T1D and appear to follow a polygenic additive model. Identifying such loci in T1D can ultimately help target who may benefit from more aggressive lipid management approaches.

2107T Association of Variants in Inflammatory Genes with Disease Severity in Familial Cerebral Cavernous Malformations Type 1. H. Choquet1, L. Pawlikowska1, J. Nelson1, C.E. McCulloch1, A. Akers2, B. Bacca2, B. Hart2, L. Morrison1, H. Kim1, Brain Vascular Malformation Consortium (BVMC). 1) Center for Cerebrovascular Research, Department of Anesthesia and Perioperative Care, University of California, San Francisco, CA; 2) Angioma Alliance, Durham, NC; 3) Department of Neurology and Pediatrics, University of New Mexico, Albuquerque, NM; 4) Department of Radiology, University of New Mexico, Albuquerque, NM. Objective: Inflammation may exacerbate disease severity in familial Cerebral Cavernous Malformations (CCM), an autosomal dominant disease characterized by multiple lesions consisting of thin leaky capillaries. The purpose of this study was to investigate whether common variants in inflammatory genes are associated with increased disease severity, as manifested by greater lesion count in familial CCM type 1 (CCM1). Methods: 178 Hispanic CCM1 subjects who all carry the same founder Common Hispanic mutation (Q455X in KRIT1 gene, named CHM) were recruited as part of the Brain Vascular Malformation Consortium (BVMC) study. Lesions were counted by a neuroradiologist on magnetic resonance imaging including susceptibility-weighted imaging performed at enrollment. Samples were genotyped on the Affymetrix Axiom Genome-Wide LAT1 Human Array. Ten candidate genes (CD14, IL1A, IL1B, IL6, IL6R, IL10, NFKB1, TGFB1, TNF and TLR4) involved in inflammatory pathways and implicated in vascular diseases or angiogenesis were selected, including 7 variants. Disease severity was quantified as residuals of log-transformed lesion count adjusting for age at enrollment and gender. Linear regression analysis was conducted for each variant assuming an additive genetic model; empirical p-values were generated by permutation to account for family structure. Results: Lesion count (range: 0 to 650.5 ± 150.5) was significantly correlated with age (R²=0.45, P<0.001). The minor alleles of variants in CD14 (rs2563298, P=0.005) and IL6R (rs114660934, P=0.007 and rs114879247, P=0.044) were associated with increased lesion count, although some of these associations did not pass Bonferroni correction for multiple testing. Variants in IL1A, IL1B, IL6, IL10, NFKB1, TGFB1, TNF and TLR4 genes were not associated with lesion count. Conclusions: Common variants in inflammatory genes CD14 and IL6R explain additional variability in lesion count independent of age and gender in familial CCM1-CHM cases; however, larger studies are needed to confirm these findings. CD14 and IL6R may have clinical importance as biomarkers of CCM disease severity and progression.

2108F Are loci associated with low density lipoprotein cholesterol (LDL-C) in non-diabetics similarly associated with LDL-C in type 1 diabetes (T1D)? S. Kurowska1,2, J. Tardif1,2, A. McLeod1,2,3, L. Surve4, R. Chagnon4,5, M. Gregoire1,2, L. Fournier6, E. Hoosie1, P.A. Cleyer7, J.M. Lachin2,3, A.D. Paterson1,2,3, DCCT/EDIC Research Group. 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 3) Department of Statistical Sciences, University of Toronto, Canada; 4) Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada; 5) Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Prosserman Centre for International Health, Toronto, Ontario, Canada; 6) Division of Cardiology, Department of Medicine, University of Toronto, Toronto, Canada; 7) The Biostatistics Center, The George Washington University, Rockville, Maryland.

Elevated LDL-C is a risk factor for cardiovascular disease, a leading cause of death in T1D. Randomization to intensive therapy, aimed at achieving glycemic control in the non-diabetic range, was associated with lower LDL-C in comparison to conventionally treated subjects in the Diabetes Control and Complications Trial (DCCT). The largest genome-wide association study of lipids to date identified 37 loci associated with LDL-C using over 95,000 primarily non-diabetic subjects (Teslovich et al., 2010). Given higher glycemia in T1D, it is unknown whether these loci are similarly associated with LDL-C in T1D. Therefore, to test this hypothesis, we used longitudinal measures of LDL-C from 1304 white participants with T1D from the DCCT. Association results from linear mixed models adjusted for time, randomized treatment group and other covariates showed that the directions for the SNP effect at 35 of the 37 loci were consistent with those reported by Teslovich et al., far exceeding the null expectation of 18.5 (p<10-7). Individual variants showed evidence of association in T1D, and correcting for multiple testing only 5 (LDLR, APOE-C1-C2, SORT1, ABCG5/A, APOB) remained statistically significant (p<0.001), suggesting a larger sample is needed to confirm the other locus-specific associations. To test the hypothesis that these 37 loci jointly contribute to susceptibility of higher LDL-C in an additive fashion, we then considered the polygenic risk score method. Risk allele was defined based on the Teslovich et al sample, and risk scores were calculated for each person by multiplying the number of risk alleles by the beta estimate from Teslovich et al for each SNP and summing the product across all 37 loci. Results of this weighted polygenic risk score analysis showed a significant additive genetic effect (p<10-38). To ensure that the result was not driven by the top 5 loci in DCCT, we repeated the analysis focusing on the remaining 32 loci. The strength of the association was attenuated as expected but remained significant (p<10-15). We also tested for a SNP by TREATMENT interaction at each locus but observed only suggestive evidence for interaction at ABO (p=0.01) and ST3GAL4 (p= 0.03). In conclusion, loci ascertained in non-diabetics are similarly associated with lipids in T1D and appear to follow a polygenic additive model. Identifying such loci in T1D can ultimately help target who may benefit from more aggressive lipid management approaches.
Identification of sequence variants of hepatic lipase (LIPC) gene in individuals with extreme HDL-C/TG levels. D. Pirm1, F.Y. Demirci1, X. Wang1, J.E. Hokenson1, R.F. Hamman2, C.H. Bunker1, M.M. Barmada1, M.I. Kambh1. 1) Human Genetics, University of Pittsburgh, GSPH, Pittsburgh, PA; 2) Epidemiology, Colorado School of Public Health, Univ Colorado Denver, Aurora, CO; 3) Epidemiology, GSPH, Univ Pittsburgh, Pittsburgh, PA.

The hepatic lipase gene (LIPC) encodes hepatic lipase (HL), which is one of the key lipolytic enzymes in HDL metabolism that regulate plasma triglyceride (TG) and HDL-C levels. Common genetic variation in lipid genes is associated with lipid variation and the risk of coronary heart disease (CHD). We hypothesized that in addition to LIPC common polymorphisms influencing plasma lipid levels, undiscovered uncommon (rare and low frequency) LIPC variants would also contribute to the lipid phenotype. In order to discover rare and low frequency variants, the entire LIPC gene (except for the very large intron 1 which is ~30 kb) was sequenced in individuals with extreme HDL-C/TG levels (95 Whites, 95 Blacks) selected from a non-Hispanic White (NHW) population from the U.S. and a Black population from Nigeria. A total of 411 variants were identified; 128 variants were shared in both populations, 53 were unique to NHWs and 230 were unique to Blacks. One hundred and eighty nine of the observed variants were novel; of which 152 were present only in Blacks and 37 were seen only in NHWs. Two novel variants were located in exons and both were synonymous; others were either intronic or located in flanking regions. In NHWs, 88 of 181 variants were common (MAF≥0.05) and 93 were low frequency or rare variants (MAF<0.05). We found 79 rare variants that are unique to either high HDL-C/low TG or low HDL-C/high TG group. In Blacks, 68 of 337 variants were common and 69 were low frequency or rare variants, of which 97 were present either in high HDL-C/low TG group or low HDL-C/high TG group. Several variants showed allele frequency differences between the two extreme HDL-C/TG groups in both populations, supporting the previous association of polymorphisms with TG and HDL-C levels. In summary, sequencing LIPC in a relatively small number of individuals with extreme HDL-C/TG levels identified a large number of novel low and rare frequency variants, with the number in blacks approximately 5x that in NHWs. Selected uncommon variants and common tag SNPs are currently being genotyped in our entire sample sets of NHWs (n=623) and Blacks (n=788) for validation and genotype-phenotype association analyses.

Replication of Hypertension Risk Score in Two Hispanic Populations. G. Beecham1, A. Beecham1, N. Vasudeva1, L. Wang1, Z. Liu1, T. Rundek1, P. Goldsmith3, M. Pericak-Vance1, D. Seo2, R.L. Sacco1,2, S. Blanton1,2. 1) John T. Macdonald Foundation Dept. of Human Genetics, Hussman Inst for Human Genomics, University of Miami, Miami, FL; 2) Department of Neurology, University of Miami, Miami, FL; 3) Division of Cardiology, University of Miami, Miami, FL.

Hypertension is an important risk factor for coronary artery disease and stroke. Although elevated blood pressure (BP) is associated with environmental factors such as diet and exercise, it is also under genetic influence. Several genome-wide association studies (GWAS) of hypertension have been conducted. In a recent meta-analysis of GWAS data from individuals of European descent the International Consortium for Blood Pressure (ICBP) has identified 29 SNPs that are genome-wide significant. A genetic risk score computed using the 29 SNPs has been replicated in Asian and African populations. Given the sparse data in Hispanics, we sought to evaluate the score computed using the 29 SNPs has been replicated in Asian and African populations. Given the sparse data in Hispanics, we sought to evaluate the effect of multiple loci was replicated (SBP p=7E-4; DBP p=2E-03), supporting the importance of these genetic variants in influencing BP in Hispanics. Of note, the effect size of the risk score in the MCR was much larger than that in the NOMAS (β=1.27 vs 0.65), with the MCR effect approaching that in European population (β=1.65) and the NOMAS effect approaching that in the African American (β=0.41). The effect size difference in the two Hispanic samples is consistent with the population substructure analysis demonstrating more European ancestry in the MCR. This suggests that, while overall the risk score replicates, the underlying population substructure is relevant to the utility of the risk score.
2113T

FLNA mutations found in patients with aortic aneurysm/dissections. H. Morisaki1, Y. Yamana1, A. Yoshida2, R. Sautana1, K. Minatoya1, Y. Shiraisa3, H. Ichikawa1, T. Kosho1, H. Sonoda1, T. Morisaki1,2,1. 1) Dept Bioscience & Genetics, NCVC Res Inst, Suita, Osaka, Japan; 2) Dept Cardiovasc Surgery, NCVC, Suita, Osaka, Japan; 3) Dept Pediatric Cardiol, NCVC, Suita, Osaka, Japan; 4) Dept Pediatric Cardiovascular Surgery, NCVC, Suita, Osaka, Japan; 5) Dept Medical Genetics, Shinshu U Grad Sch Med, Matsumoto, Nagano, Japan; 6) Dept Cardiovascular Surgery, Kyushu U Grad Sch Med, Fukuoka, Fukuoka, Japan; 7) Dept Mol Pathophysiol, Osaka U Grad Sch Pharm Sci, Suita, Osaka, Japan.

Thoracic aortic aneurysm/dissections (TAAD) often result from genetic predisposition, since about 10 to 20 % of these patients have a first degree relative with TAAD. To find disease-causing gene mutations for TAAD, we generated exome sequence from 213 individuals with young-onset or familial TAAD and/or related connective tissue disorders, who were confirmed not to have any pathogenic mutation in FBN1, TGFBR1, TGFBR2, ACTA2 or SMAD3. Sequence data was interrogated for sequence changes, compared with the data base of 1000 Genomes. While pathogenic mutations in several genes potentially related to aortopathy emerged from this screening, FLNA mutations were recurrently identified in 4 probands. Three females (20-54 years old) were heterozygotes for nonsense or frame-shift mutations and 1 male (22 years old) was a hemizygote for missense mutation in the X-linked FLNA gene, encoding filament A. Family study revealed a mother of the male patient carried the same FLNA mutation as her son, and she also had dilated cardiomyopathy with severe mitral valve insufficiency. Brain MRI study exhibited periventricular nodular heterotopias (PNH) in all 5 patients. Further studies on the clinical features showed several cardiovascular events including aortic surgery in 2, severe mitral valve insufficiency in 3, aortic valve insufficiency ins in 3, severe orthostatic hypotension in 2 and patent ductus arteriosus in 1. Also, several features related to connective tissue disorders, including joint hypermobility in 4, inguinal hernia in 3, thoracic deformity in 2, and external entropion in 2, were observed. Regarding neurological features, all had normal intelligence although both patients were diagnosed as adult-onset epilepsy. FLNA was shown as a responsible gene for X-linked PNH. Also, several distinct genetic disorders, including opotopalatodigital syndrome, frontometaphyseal dysplasia, Melnick-Needles syndrome and Ehlers-Danlos syndrome, were caused by FLNA mutation. The major clinical features of our cases were aortic disorder with some connective tissue involvement. Based on these results and literature reviews, we suggest that it is advisable to examine FLNA mutation in cases with cardiac involvement. We propose to screen cardiovascular manifestations in individuals with X-linked PNH.

2115W

A Variant in the 5-Hydroxytryptamine Receptor 1A (HTR1A) Gene Is Associated with Platelet Reactivity and Incident Cardiovascular Events. S.H. Shah1,2, M.A. Babyak2, E.R. Hauser1, D. Craig1, M. Christ-Ladd1, C. Haynes3, B. Brunnett3, W.E. Kraus1,2, R. Becker1, R.B. Williams2.

1) Division of Cardiology, Department of Medicine, Duke University Medical Center, Durham, NC; 2) Duke Institute for Molecular Physiology, Durham, NC; 3) Department of Psychology and Behavioral Sciences, Duke University Medical Center, Durham, NC.

Background. Variation in the 5-hydroxytryptamine receptor 1A (HTR1A) gene has been associated with psychiatric phenotypes. Given that depression has been linked to cardiovascular disease (CVD) events, we hypothesized that HTR1A SNP rs1364043 is associated with platelet reactivity (an intermediate phenotype for CVD events), and is associated with CVD event endpoints. Methods. Rs1364043 in HTR1A was genotyped using an ABI TaqMan assay in 8697 participants in the CATHERGENIO study of patients referred to Duke University for cardiac catheterization for evaluation of CVD. Incident events were defined as all-cause death and/or myocardial infarction (MI) after enrollment. Race-stratified Cox proportional hazards modeling was used to test for association with time-to-event in two models (death and death or MI), adjusted for age, sex, hypertension, diabetes, ejection fraction, CAD, dyslipidemia, smoking, and BMI. Expression quantitative trait loci (eQTL) analyses were used to test for cis and trans effects of SNP on RNA expression using additive multivariable linear regression. Association between rs1364043 and platelet reactivity was tested in N= 410 healthy individuals, defined as a binary trait of hyperreactivity (>60% aggregation to epinephrine [2 μM] and serotonin [10 μM]), using generalized linear models adjusted for age, sex and race. Results. In whites, the C allele of rs1364043 was associated with increased platelet reactivity (HR 1.09, 95% CI [1.02-1.12], p<0.04) and dominant models (HR 1.11 [1.001-1.24], p=0.05), and with time to death/MI (HR 1.10 [1.02-1.19], p=0.02 for additive and HR 1.13 [1.03-1.25], p=0.02 for dominant model). eQTL analyses identified no cis effects. The top transcripts associated with rs1364043 including several genes involved in cell cycle processes, protein synthesis and stability including CDK11a, RPL32, ASNS, DDHD2, MED27 and STOML2 (p=0.002-0.0003). Platelet reactivity studies showed a trend for increased proportion of individuals with high platelet reactivity in rs1364043 C allele homozygotes as compared with heterozygotes or wildtype (73.7% vs. 58.0% vs. 52.0%, p=0.04, dominant model). Conclusions. Genetic variants associated with intermediate traits may themselves serve as disease markers for those phenotypes. Studies with intermediate psychosocial traits, is also associated with platelet hyperreactivity and predicts CVD events, with an up to 13% increased risk.

2116T


1) Adult Med Gen Prog, University of Colorado - CU Cardiovascular Institute, Aurora, CO; 2) Temple University School of Medicine, Philadelphia, PA.

Background: Recently BAG3 has been identified as a cause of genetic forms of dilated cardiomyopathy. The consequences of BAG3 mutations on cardiac pathology and the role of BAG3 in heart failure are not known. Methods: Whole exome DNA analysis was performed on a multi-generational family affected with dilated cardiomyopathy. Prior studies including linkage analysis, candidate gene analysis, and SNP array analysis had failed to locate the pathogenic mutation. Exome capture was achieved using the Agilent SureSelect Exome capture system and sequencing was completed on an Illumina HiSeq2000. Analysis of explanted human heart tissue from one mutation carrier was performed and BAG3 levels in control and failing hearts were assayed by Western blot analysis. Results and Conclusions: A 10-nucleotide deletion in exon 4 of the BAG3 gene was identified through exome sequencing from distantly affected family members. The mutation was present in seven affected family members and absent in eight healthy relatives. Standard histology of a BAG3 mutation carrier’s explanted heart revealed mild to moderate myocyte hypertrophy and patchy mild to moderate interstitial fibrosis; changes more prominent in the left ventricle and septum than the right ventricle. Also, electron microscopy suggested increased glycogen deposition, possibly related to the role of BAG3 protein in autophagy and cellular trafficking. BAG3 levels were markedly depressed with evidence of truncated BAG3 protein in the explanted heart specimen, consistent with haploinsufficiency. Conclusions: Exome sequencing is increasingly useful in identifying pathogenic mutations in cardiomyopathy families where distant relatives are available for segregation analysis. The microarray of BAG3 heart failure haploinsufficiency related to apparent haploinsufficiency of BAG3, suggesting that modulation of BAG3 expression could represent a novel avenue for therapeutic exploration.
2117F
Haploinsufficiency of RERE contributes to cardiovascular defects associated with 1p36 deletions. H. Zaveri1, B. Kim2, A. Hernandez-Garcia2, T.F. Beck2, D.A. Schelochkov2, M. Justice1, B. Lee1, S.R. Lalani1, D.A. Scott1
1) Molec & Human Gen, Baylor College Med, Houston, TX; 2) Dept of Pediatrics, The University of Iowa, Iowa City, IA.
Terminal deletions of chromosome 1p36 affect 1 in 5000 newborns and cause a clinically recognizable syndrome characterized by cognitive impairment, eye/vision problems, hearing loss, facial clefting, cardiovascular malformations and cardiomyopathy. Although cardiovascular anomalies and/or cardiomyopathy affect 71% of individuals with 1p36 deletions, and are the most common cause of morbidity and mortality in the newborn stage, the genes that contribute to these phenotypes have yet to be identified. Using a clinical and molecular cytogenetic data from individuals with terminal and interstitial deletions, we have created a deletion/phenotype map of the 1p36 region. This map includes five non-overlapping critical regions for cardiovascular defects and two non-overlapping critical regions for cardiomyopathy. The majority of these regions contain positional candidate genes which have been shown to cause cardiovascular malformation and/or cardiomyopathy in mice. These genes include DVL1, SKI, PRDM16, UBE4B, FDPN, SPEN, PDPN, HSPG2, CDC42, and LUZP1. The arginine-glutamic acid dipeptide repeats gene (RERE) encodes a nuclear receptor coregulator and is located in a critical region for both cardiovascular malformations and cardiomyopathy on 1p36 and is required for normal retinoic acid signaling. Since perturbations in retinoic acid signaling can cause cardiovascular malformations, and retinoic acid signaling suppresses myocardial hypertrophy and cardiac fibrosis, we hypothesized that RERE deficiency contributes to the cardiac phenotypes seen in individuals with terminal and interstitial 1p36 deletions that include RERE. To test this hypothesis, we generated an allelic series of RERE-. seen in individuals with terminal and interstitial 1p36 deletions that include RERE. Cardiac defects were also documented in mice in which RERE was conditionally ablated in the first and second heart fields. We conclude that RERE functions in a cell autonomous manner to direct cardiac development and that deletion of RERE contributes to the cardiac malformations and cardiomyopathy associated with 1p36 deletions.

2118W
Analysis of Common and Coding Variants with Cardiovascular Disease in the Diabetes Heart Study. J.N. Adams1,2,3, L.M. Raffield1,2,3, D.I. Freeman4, C.D. Langefield5, M.C.Y. Ng2,3, J.J. Carr6, A.J. Cox3,7, D.W. Bowd- en1,2,3,7
1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Internal Medicine - Nephrology, Wake Forest School of Medicine, Winston-Salem, NC; 5) Division of Public Health Sciences, Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 6) Department of Radiologic Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 7) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC.
Type 2 diabetes (T2D) affects ~8% of the US population. T2D is a major risk factor for cardiovascular disease (CVD). Up to 65% of T2D deaths are CVD-related. Identification of genetic risk factors for CVD is important to understand individual disease risk. Recent GWAS meta-analyses in the CHARGE consortium identified loci associated with coronary artery calcification (CAC), myocardial infarction (MI), carotid intima media thickness (IMT), and/or carotid plaque. CAC and IMT are clinically relevant measures of subclinical CVD and strongly predict mortality. We examined if prior associations from the population-based CHARGE study were applicable in the Diabetes Heart Study (DHS), a family-based cohort (n=1220) with >80% T2D affected subjects. SNPs (n=36) from CHARGE were evaluated in GWAS data from the DHS. Phenotypes including vascular calcification and conventional CVD risk factors were tested. No SNPs were significantly associated after correction for multiple comparisons (p=0.0014), although multiple SNPs showed nominal significance: rs999839 in PIK3G (p=0.0315, p=0.0084) and rs646776 in CELSR2 (p=0.380, p=0.0099) were associated with CAC; rs17398575 in PIK3G was associated with aortic calcification (AACP) (p=11.0, p=0.054). Additional SNPs at COL4A2 and CXCL12 were nominally associated with all-cause or CVD-mortality. Genes implicated by CHARGE were investigated further by examining exonic variants. 209 coding variants were investigated using illumina HumanExome BeadChip genotype data from the DHS; ns61730407 in OR2G3 was significantly associated with T2D (p=0.026; OR=1.15). We found that genetic risk factors for subclinical CVD in the general population (CHARGE) were, in part, associated with T2D-related risk factors and outcomes.
**2119T**

**VKORC1** and CALU eQTLs and susceptibility to venous thromboembolism among African Americans. W. Hernandez1, E.R. Gamazon1, A. Konkashbaev1, K. Aquino-Michaels1, T.J. O’Brien2, A.P. Harralson3, R.A. Kittles4, A. Barbour4, M. Tuck5, S.D. McIntosh4,5, J.N. Douglas6, L.H. Cavallari6, M.A. Perera1. 1) Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL; 2) The George Washington University, Department of Pharmacology and Physiology Washington, DC; 3) University of Illinois, Department of Medicine, Institute of Human Genetics, Chicago, IL; 4) The George Washington University, Department of Medicine, Washington, DC; 5) Uniformed Services University of the Health Sciences, Department of Veterans Affairs, Washington, DC; 6) University of Illinois, Chicago, Department of Pharmacy, Chicago, IL.

Venous thromboembolism (VTE) is a complex disease encompassing deep vein thrombosis (DVT), pulmonary embolism (PE), or a combination of both. In the US, each year over 900,000 individuals are hospitalized due to VTE and over 300,000 patients die. African Americans (AAs) have the highest incidence and mortality rates of DVT and PE. Warfarin is used to treat and prevent DVT and/or PE. However, warfarin dose requirement has been shown to be higher among AAs as well as DVT/PE patients. In this study, we aimed to investigate the role between VKORC1 and CALU expression quantitative trait loci (eQTLs) and DVT/PE susceptibility. As subtle differences in plasma levels of proteins involved in the coagulation pathway may be risk factors, eQTLs may yield novel insights into the genetic risk of DVT/PE. Using publicly available liver eQTL data, we identified both cis and trans eQTLs for VKORC1 and CALU genes known to be associated with warfarin dose. 72 SNP eQTLs were genotyped in a study population of 462 AAs on stable warfarin dose; of which 256 individuals were treated with warfarin due to DVT/PE (cases) and the remaining due to a variety of other conditions including atrial fibrillation (controls). We found a significant decrease in risk of DVT/PE for carriers of the minor allele of two VKORC1 eQTLs and one CALU eQTL (rs9925964, OR=0.53, P=0.01; rs12597511, OR=0.48, P=0.02; and rs11054879, OR=0.61, P=0.03 respectively). We also found our DVT/PE patients had a higher percentage of West African ancestry and were younger compared to controls (t=1.991, P=0.04 and t=2.720, P=0.007 respectively). Interestingly, the frequency of these protective alleles were approximately 10% lower in our study population and significantly higher in the HapMap European Americans (CEU) at approximately 40% but lower among Yorubans (YRI) at 5%. The lower frequency of these protective alleles among populations of African descent compared to European Americans (EAs) is particularly interesting as the risk of DVT/PE is lower among EAs than AAs. In this study, we also observed VKORC1 -1639 is highly associated with warfarin dose requirement (t=-0.42 and P=6.79E-12) but not to DVT/PE susceptibility (OR=0.79 and P=0.32). By investigating eQTLs for genes known to contribute to warfarin dose requirement we have uncovered novel disease loci involved in the coagulation cascade and risk of DVT/PE. These findings may help explain the increased risk of DVT/PE seen in the African American population.

**2120F**

Multi-allelic Haplotype Association Identifies a new Protective Haplotype in the KCNN3 Gene for Post-Operative Atrial Fibrillation. M. Heydarpoor1, C. Collard2, A. Fox1, J. Muehlischlegel1, M. Sigurdsson1, S. Sharma1, S. Body1. 1) Department of Anesthesiology, Preoperative and Pain Medicine, Brigham and Women’s Hospital, Boston, MA; 2) Division of Cardiovascular Anesthesia, Texas Heart Institute, Houston, TX.

Atrial fibrillation (AF) is the most common arrhythmia after cardiac surgery. Two regions have been associated with AF in ambulatory populations (KCNN3, ZFHX3), but not yet associated with AF in cardiac surgical populations. Candidate single nucleotide polymorphisms (SNPs) that describe risk haplotypes for AF in ambulatory populations were genotyped in this study. The objective was to replicate findings of AF-associated loci while accounting for previous association of 4q25 with postoperative AF (poAF). Using a case-control association study of 325 Caucasian patients with poAF and 727 Caucasian controls, we examined 488 SNPs in three regions (1q21.3, 4q25 and 16q22.3) for SNP and haplotype association analysis using logistic regression analysis while accounting for age, gender, prior AF and other clinical variables, and an additive genetic model. Maximum likelihood hierarchical clustering methods were used for haplotype clustering analysis. Permutation testing was used to assess statistical significance. Significant association was identified between SNPs rs6683557 (1q21.3, OR=0.70, P=0.00051), rs3853445 (4q25, OR=0.68, P=0.0016), and rs67402452 (16q22.3, OR=1.45, P=0.0033) and poAF after adjusting for prior association of SNPs in 4q25 and for clinical variables. We identified a 1q21.3 haplotype, AAAAA which conferred significant protection against poAF (OR=0.56, P=0.0011). All five SNPs in this haplotype were intronic to KCNN3, which encodes a potassium channel protein involved in atrial repolarization. The rs3853445 in 4q25 region as a protective variant against AF has been reported by Lubitz et al., 2010, (OR=0.71, P=4.1×10-5) which is confirmed by this study. Two other susceptible markers (rs2200733 & rs13143308) are identified in 4q25 for risk of poAF which previously reported by Body et al. 2009 to predict poAF after coronary artery bypass graft (CABG). These results suggest that a common haplotype AAAAA of KCNN3 gene confers a significant protective effect on the development of poAF. Our results confirmed the role of these two regions variants for AF risk in previous studies (Body et al. 2009, Body et al. 2009, Gregory et al. 2012) for European-descent population. Consideration of multiple susceptibility signals at these regions identifies individuals in high risk of AF or protective against AF and they may localize regulatory elements at the locus with biological relevance in the pathogenesis of AF.
2121W
Elucidating the role of genes encoding sarcomeric structural proteins in modulating cardiac hypertrophy in Hypertrophic Cardiomyopathy. C.J. Kinnear1, L. Bloom1, L. Van der Merwe1, T. Reevera2, M. Heradien1, A. Goosen1, P. Brink2, J.C. Moolman-Smook1.
1) University of Stellenbosch/ Medical Research Council (US/MRC) Centre for Molecular and Cellular Biology, Department of Biomedical Sciences, Stellenbosch University, South Africa; 2) Biodiagnostics Unit, Medical Research Council of South Africa; 3) Department of Cardiology, IRCSS San Matteo Hospital, Pavia, Italy; 4) Department of Medicine, Faculty of Health Sciences, University of Stellenbosch, South Africa.

Hypertrophic cardiomyopathy (HCM) is an autosomal dominantly inherited cardiac muscle disease clinically defined by unexplained left ventricular hypertrophy (LVH). The degree and location of the LVH in HCM patients is extremely variable, even in patients carrying the same HCM-causing mutations. Phenotypic expression of HCM is therefore dependent on the disease-causing mutation as well as additional determinants including environmental factors and modifier genes. Many investigations aimed at identifying modifier genes in HCM have focused on genes that are relatively far removed from the primary disease-causing genes. We hypothesise that since most HCM-causing mutations are found in genes encoding proteins of the cardiac sarcomere, factors that closely interact with defective sarcomeric proteins, will have a greater capacity to modulate them and thus modulate the LVH phenotype. Therefore genes encoding sarcomeric structural proteins as well as those encoding enzymes that function in sarcomere-based energetics may be considered plausible modifier genes of LVH in HCM. In the present study we focused on genes encoding sarcomeric structural proteins as candidate modifiers of LVH. A total of 24 single nucleotide polymorphisms (SNPs) in 15 genes (ACTC1, TNNT2, TNIN3, TPM1, ACTN3, MYL3, MYH7, MYH2, MYL2 and MyBCP3) were investigated for association with cardiac hypertrophy traits in a familial HCM-cohort. A total of 388 individuals from 27 HCM South African HCM families in which one of three HCM founder mutations (R95WNTN1, R460WMYH7 and A797MYH7) were genotyped by validated Taqman® SNP genotyping assays. All participants provided written informed consent and was conducted in accordance with the ethical guidelines as set out in the Declaration of Helsinki, 2008. We identified a number of loci which were significantly associated with phenotypic expression of LVH, as determined based on disease-phenotype correlations in HCM and could potentially improve patient risk stratification and management.

2122T
Analysis of Coding Variants in C1q/TNF Superfamily Genes in the Diabetes Heart Study. L.M. Raffield1,2,3, A.J. Cox2,3,4, C.D. Langefeld2, M. Hara2,1, T. Xie2,1, A. Freundman2,1, V. X. Friederman2,1, D.W. Goosen1,2,1, J.C. Moolman-Smook1.
1) University of Stellenbosch/ Medical Research Council (US/MRC) Centre for Molecular and Cellular Biology, Department of Biomedical Sciences, Stellenbosch University, South Africa; 2) Center for Human Genomics, Wake Forest University Health Sciences, Winston-Salem, NC, USA; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA; 5) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA; 6) Department of Radiologic Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA; 7) Department of Internal Medicine - Nephrology, Wake Forest School of Medicine, Winston-Salem, NC, USA.

Members of the C1q/tumor necrosis factor (TNF) superfamily have a variety of roles in inflammation and development of cardiometabolic and have been associated with cardiovascular disease (CVD) and type 2 diabetes (T2D). Previously, we have shown that coding variants in the collagen domain of the C1q family member adiponectin dramatically lower serum adiponectin levels and impede multimerization. We hypothesized that coding variants influencing multimerization or other protein-protein interactions in C1q/TNF superfamily members and their binding partners/receptors could impact CVD- and T2D-related phenotypes. This analysis was performed in 1190 European American individuals from 468 families in the Diabetes Heart Study (DHS). In the DHS cohort, 83.7% of participants were affected by T2D with average disease duration of 10.5 ± 7.2 years. Genotyping was completed using the Illumina Infinium HumanExome BeadChip array. 464 coding variants in 97 C1q/TNF related genes were analyzed for associations with CVD- and T2D-related phenotypes, including fasting glucose and 2-h post an oral glucose tolerance test (P = 9.11 × 10\(^{-5}\)). Replication of these uncommon coding variants in additional cohorts will be necessary, but these results suggest important roles for C1q/TNF superfamily members in CVD- and T2D-related phenotypes which can be revealed by analysis of exonic variants.

2123F
Cross-sectional and Longitudinal Replication Analyses of Genome-wide Association Loci of Type 2 Diabetes in Han Chinese Populations. Q. Zhao1, X. Kong2, J. Hong3, X. Zhang4, J. He1, W. Yang2.
1) Department of Epidemiology, Tulane University School of Public Health, New Orleans, LA; 2) Department of Endocrinology, China-Japan Friendship Hospital, Beijing, China.

Most of the genetic loci for type 2 diabetes (T2D) identified by genome-wide association studies (GWAS) in populations of European ancestry are still controversial about their associations with T2D of East Asians. The purpose of our study is to examine recently GWAS-identified loci for their associations with T2D and quantitative glycomic traits and their effects on longitudinal changes in fasting plasma glucose (FPG) and T2D development in Han Chinese populations. Single nucleotide polymorphisms (SNP) from 26 loci were genotyped in 10,011 Chinese Hans (5,338 T2D cases and 4,663 controls). Eight SNPs in or near CDKN2A/2B (OR = 1.11 × 10\(^{-5}\)), were also associated with glucose level of 2-h post an oral glucose tolerance test (P = 9.91 × 10\(^{-5}\)) and insulinogenic index (P = 2.71 × 10\(^{-5}\)). A total of 16 SNPs showed consistent association directions with the reported T2D GWAS meta-analysis results of the Asian Genetic Epidemiology Network (AGEN) (6,952 cases and 11,865 controls). Twelve SNPs of the 16 aforementioned SNPs reached the significance of 0.05 in a combined analysis of AGEN and the current study (including 24,818 subjects), and most of them had similar effect sizes to those seen in European populations. In a cohort of 1,881 Chinese Hans without T2D at baseline, individuals carrying more risk alleles of the replicated SNPs in the combined analysis had greater FPG increase and T2D incidence over a 7.5-year follow-up period, with each quartile increase in the number of risk alleles being associated with a 0.06 mmol/l greater increase in FPG (P = 0.03) and 19%0 higher odds for developing T2D (P = 0.058). Our results indicate that some of GWAS-identified loci for T2D in Europeans are associated with T2D and glycomic traits and may also predict risk for T2D development in the Chinese population.
2124W
Sex-specific genetic variants on 11p15 influence high density-lipoprotein cholesterol levels in long-lived subjects: The Long Life Study Family (LLFS). M.F. Feitosa1, M. Vinggaard2, C. Shungu1, M. Kammerman1, D. Shungu1, R.J. Straka3, J.H. Lee4, K. Christensen5, A.B. Newman6, M.A. Province7, I.B. Borcik8, 1) Department of Genetics, Washington University School of Medicine, St. Louis, MO; 2) Department of Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 3) Department of Experimental and Clinical Pharmacology, College of Pharmacy, University of Minnesota, Minneapolis, MN; 4) Brower Laboratory and linkage analysis was carried out using SOLAR. For subjects taking lowering-lipid medications, a constant was added to their measured HDL-C levels specific to the reported class of medication. HDL-C was then adjusted to remove these effects prior to principal component analysis of stepwise regression, within sex. Significant linkage was found on 11p15 for HDL-C in women (LOD=4.2) at 19 cM, but not in men (LOD=0.5) or in sex-combined data (LOD=2.2). To test whether common variants from GWA explained this linkage peak, we further adjusted HDL-C for all 30 SNPs in our GWA SNPs with p < 1e-4 and repeated the linkage analysis. Despite the decrease in LOD from 4.2 to 3.0, evidence of linkage remained, suggesting that other variants in the region influence HDL-C, and these may be rare. Because it is likely that different variants are segregating in different families, we aggregated the 277 families (1,065 women) supporting linkage to this region with Heterogeneity LOD=20.9. The strong evidence of genetic effects in these families were within 5-11 Mb region. Chromosome regions thus detected can be prioritized for sequencing to identify the specific rare variants influencing HDL-C in these healthy-longevity families.

2125T
Influence of genetic determinants and lifestyle factors on blood lipid levels in long-lived subjects: The Long Life Study Family (LLFS). M.F. Feitosa1, M. Vinggaard2, C. Shungu1, M. Kammerman1, D. Shungu1, R.J. Straka3, J.H. Lee4, K. Christensen5, A.B. Newman6, M.A. Province7, I.B. Borcik8, 1) Department of Genetics, Washington University School of Medicine, St. Louis, MO; 2) Department of Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 3) Department of Experimental and Clinical Pharmacology, College of Pharmacy, University of Minnesota, Minneapolis, MN; 4) Brower Laboratory and linkage analysis was carried out using SOLAR. For subjects taking lowering-lipid medications, a constant was added to their measured HDL-C levels specific to the reported class of medication. HDL-C was then adjusted to remove these effects prior to principal component analysis of stepwise regression, within sex. Significant linkage was found on 11p15 for HDL-C in women (LOD=4.2) at 19 cM, but not in men (LOD=0.5) or in sex-combined data (LOD=2.2). To test whether common variants from GWA explained this linkage peak, we further adjusted HDL-C for all 30 SNPs in our GWA SNPs with p < 1e-4 and repeated the linkage analysis. Despite the decrease in LOD from 4.2 to 3.0, evidence of linkage remained, suggesting that other variants in the region influence HDL-C, and these may be rare. Because it is likely that different variants are segregating in different families, we aggregated the 277 families (1,065 women) supporting linkage to this region with Heterogeneity LOD=20.9. The strong evidence of genetic effects in these families were within 5-11 Mb region. Chromosome regions thus detected can be prioritized for sequencing to identify the specific rare variants influencing HDL-C in these healthy-longevity families.

2126F
Multi-ethnic case-control studies of Moyamoya disease by high-depth exome sequencing of unrelated affected individuals and matched controls. M. Clark1, L. Shoemaker2, A. Patwardhan1, R. Chen1, G. Chandrati-lake1, S. Garcia1, N. Leng1, S. Chervitz1, M. Pratt1, H. Lam1, D. Newburger1, S. Kirk1, C. Haudenschild1, J. West1, R. Chen1, G. Steinberg3, 1) Personalis, Inc., Menlo Park, CA; 2) Department of Neurosurgery, Stanford University School of Medicine, Stanford, CA; 3) Stanford Institute for Neuro-Innovation & Translational Neurosciences, Stanford University, Stanford, CA. Moyamoya disease (MMD) is a rare cerebrovascular disorder characterized by stenosis/occlusion of the internal carotid artery and proliferation of microvessels. Symptoms include stroke, seizures, aphasia and cognitive impairments. It exhibits low penetrance autosomal dominant inheritance. To determine genetic associations with MMD, 125 ethnically diverse unrelated patients were matched based on gender and broad ethnic categories to 125 controls obtained from the 1000 Genomes Project. All case and control samples were exome sequenced to high depth and analyzed together using the Personalis Pipeline. Rather than using public 1000 Genomes data, case and control exomes were resequenced on the same platform and analyzed together. Case-control Genotype calls were performed within broad ethnic subsets (70 Asian, 136 Caucasian) to identify enriched variants. In order to investigate associations with MMD among cases in which the known founder RNF213 mutation was absent, an additional matched case-control analysis using only samples without the mutation was conducted. Variants were collapsed into genes to identify genes enriched for mutations in these MMD cases. The most highly enriched variant in Asian cases, R4810K (p=6.01×10-5), confirmed a known founder mutation in the gene RNF213 in the Asian population. Notably, this mutation was highly enriched in East Asians, but not Southeast Asian or Pacific Islanders (p=9.52×10-4). A second variant (R4810G) was identified at the same locus as the East Asian founder mutation in one self-identified Filipino sample. Fine-mapping, nonsense, and splice site mutations were not observed. SNPs RNF213 was not enriched in the case and control samples, suggesting that severe loss of function mutations in RNF213 may not be tolerated. In the Caucasian subset, the most highly enriched variant was P562L in the ZDC gene (p=7.93×10-4), but no strong founder mutation was identified. Among samples without the mutation, the exome capture method was ranked OBSCN as the gene most enriched for variants (p=5.31×10-5). Our analyses leveraged high-depth exome sequencing, an internally re-sequenced multi-ethnic control set, and an analysis pipeline that identified disease associations despite limited sample sizes. We independently confirmed the RNF213 founder mutation for MMD in the East Asian population, found it absent in other ethnicities, and identified new candidate genes via collapsing methods.
2128T

Digenic inheritance of mutations in ITGA7 and MYH7B results in congenital myopathy with left ventricular non compact cardiomyopathy. T. Esposito², S. Sampolo², G. Limongelli², A. Varone³, D. Formico³, D. Di Diodato⁴, O. Farina⁴, F. Napolitano⁴, G. Pacileo⁴, F. Gianfrancesco¹, G. Di Iorio². 1) Inst Gen & Biophysics, Italian Natl Res Council, Naples, Italy; 2) Department of Medical Sciences, Surgery, Neurological, Metabolic and Aging, Second University of Naples, Italy; 3) Department of Cardiological Sciences, Second University of Naples, Italy; 4) Department of Neuro-sciences, Santobono-Pausilipon’ Hospital, Naples, Italy.

We describe an unusual association between congenital fiber type disproportionate (CFTD) and left ventricular non compaction cardiomyopathy (LVNC) in the proband of an Italian family. The CFTD myopathy is a genetically heterogeneous disorder characterized by relative hypertrophy of type 1 muscle fibers compared to type 2 fibers on skeletal muscle biopsy. The hallmark features of LVNC include prominent trabeculations and deep endocardial recesses associated with arrhythmias, thromboembolic events, and heart failure. Clinical and instrumental analysis of the family members identified the LVNC phenotype also in the mother, the sister and the first-degree cousin of the proband suggesting a dominant mode of inheritance with high phenotype heterogeneity and reduced penetrance. Recessive mode of inheritance was assumed for the CFTD phenotype which is characteristic of the proband. The study was focused at the identification of the responsible gene/s through whole exome sequencing approach. Two homozygous missense mutations in two genes, the myosin heavy chain 7B (MYH7B) and the integrin alpha 7 (ITGA7) were identified. Both genes are expressed in heart and muscle tissues and both mutations were predicted as deleterious and were not found in healthy population. The R890C mutation in the MYH7B gene segregates with the LVNC phenotype in the examined family, and was also found in one not related patient affected by LVNC. The arginine at position 890 of the MYH7B gene is highly conserved in all species and this region is also conserved in MYH7 gene in which when mutated causes cardiomyopathy. The glutamic acid at position 882 is highly conserved in all species and bioinformatics tools predict that when the 882E is substituted with the lysine the secondary structure of the protein changes. This study identifies two novel disease genetic variants in two independent loci.

2129F

Genetic variants associated with C-reactive protein in African Americans: a MetaboChip analysis in the Population Architecture using Genomics and Epidemiology (PAGE) study. J. M. Kocarnik¹, C. L. Carly², A. Reiner¹, C. L. Avery², C. M. Ballantyne³, J. Haessler⁴, A. LaCroix⁵, A. Young⁶, T. L. Assimes⁶, M. Barbacid⁶, W. Tang⁶, L. G. Best⁷, J. L. Ambite⁸, I. Cheng⁹, L. A. Hindorff¹⁰, G. Heiss¹¹, C. A. Haiman¹¹, C. L. Kooperberg¹, L. Le Marchand¹², U. Peters¹. 1) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 5) Division of Epidemiology, Human Genetics & Environmental Sciences, The University of Texas, Houston, TX; 6) Division of Epidemiology & Community Health, University of Minnesota, Minneapolis, MN; 7) Missouri Breaks Industries Research Inc., Timber Lake, SD; 8) Information Sciences Institute, University of Southern California, Marina del Rey, CA; 9) Cancer Prevention Institute of California, Fremont, CA; 10) Division of Genomic Medicine, NHGRI, NIH, Bethesda, MD; 11) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 12) University of Hawaii Cancer Center, Honolulu, HI.

Introduction: Inflammation is an important health outcome related to many common complex diseases, several of which have differential disease burden by ethnicity. C-reactive protein (CRP) is a circulating biomarker indicative of systemic inflammation. Genome-wide association studies (GWAS) have successfully identified susceptibility loci important to inflammation and inflammation-related diseases, though discovered primarily in populations of European ancestry. To test if genetic variants associated with inflammation are generalizable to populations of African ancestry, we evaluated the association between serum CRP levels and the large number of single nucleotide polymorphisms (SNPs) on the MetaboChip array. Methods: We analyzed 9,648 African American participants from three studies—Atherosclerosis Risk in Communities (ARIC), Multiethnic Cohort (MEC), and Women’s Health Initiative (WHI)—participating in the Population Architecture using Genomics and Epidemiology (PAGE) study or the SNP Health Association Resource (SHARE) project. Participants had serum high-sensitivity CRP measure-ments and genotype information for the 196,725 SNPs on the MetaboChip, a customized Illumina iSelect array targeting 257 regions identified by GWAS for metabolic, atherosclerotic and cardiovascular endpoints. Fixed-effect meta-analyses combined study-specific linear regression estimates to evaluate the association between each SNP and log-transformed CRP. Additive genetic models adjusted for age, sex, top 4 principal components of genetic ancestry, and study-specific factors. Results: 28 SNPs in 5 regions reached a Bonferroni-corrected p-value cutoff of p<3.3e-7. CRP (10 SNPs below the correction threshold; rs3091244 had the lowest p-value at p=4.2e-76), TOMM40-APOE-APOC1 (8 SNPs; rs446037 p=2.9e-16), NOS1AP-OLFML2B (5 SNPs; rs6676438 p=3.7e-10), HNF1A (4 SNPs; rs7979473 p=4.8e-10), and LEPR (rs4655779 p=2.6e-7). Discussion: We identified five genetic regions that were associated with CRP levels in an African American population. One of these regions, NOS1AP-OLFML2B, has not previously been associated with CRP and may represent a novel risk locus for CRP levels. SNPs in this region were included on the MetaboChip because of previous associations with QT interval. Future analyses will further evaluate these associations in other race/ethnicity groups, compare LD patterns in these regions between these groups, and use conditional analyses to identify independent loci.
2130W
Predictive Profile for the early detection of metabolic syndrome in a pediatric Turkish cohort. M.J. White1,2, M. Agirbasli2, F. Ezrin1, D. Agirbasli2, S.M. Williams1,2 1) Center for Human Genetics Research, Vanderbilt Univ, Nashville, TN; 2) Department of Genetics, Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH; 3) Department of Cardiology, Marmara University School of Medicine, Istanbul, Turkey; 4) Department of Medical Biology, Acibadem University School of Medicine, Istanbul, Turkey; 5) Department of Medical Biology, Acibadem University, Istanbul, Turkey.

Metabolic syndrome (MetS) is characterized by the presence of three or more cardiovascular disease (CVD) risk factors, including abnormalities in lipid profiles. The presence of MetS increases CVD risk by approximately two-fold and is also associated with increased mortality in adults. Recent evidence of childhood MetS underscores the impact of the global childhood obesity burden and its potential impact on long-term health. MetS prevalence rates are as high as 60 percent in some overweight pediatric/adolescent populations and autopsy studies in children and adolescents have revealed that the presence of MetS is related to the early stages of atherosclerosis. These data indicate that CVD morbidity and mortality may be increased through childhood MetS. We investigated the impact of genetic variation (single nucleotide polymorphisms - SNPs) in six candidate genes known to influence lipid profiles and circulating sex hormone levels in adults, as well as known CVD risk factors, on susceptibility to MetS in a pediatric Turkish cohort (n=360). Turkish children and adolescents have lipid profiles characterized by high-density lipoprotein cholesterol, elevated triglyceride levels (two hallmarks of MetS), and lower prevalence of obesity than American pediatric populations. This unique lipid profile makes Turkish children and adolescents an at-risk population for MetS. Logistic regression analysis revealed that a SNP in SHBG (rs1799941) was significantly associated with MetS (p=0.010) after adjustment for known CVD risk factors, suggesting an independent genetic effect. Exploratory single and multi-focus analysis was performed using multi-factor dimension reduction (MFR) to determine the genetic and metabolic profile most predictive of MetS. MDR identified rs1799941, high triglycerides, BMI, and insulin resistance as the most predictive model (Testing balance accuracy=0.7965, p<0.001). In conclusion, our study identified a potential SNP in SHBG (rs1799941) as a genetic risk factor for MetS in a pediatric population. This result has the potential for future clinical use.

2131T
Investigation of functional variants of eight SNPs in lipid level modifier genes in healthy Roma and Hungarian population. B. Melegh1,2, K. Susemi1,2, L. Jaromi1,2, L. Magyari1,2, E. Kovess1,2, B. Duga1, R. Szalai1, P. Matyas1, Zs. Bantar1, A. Szabo1, J. Bene1. 1) Department of Medical Genetics, University of Pecs, Pecs, Hungary; 2) Szentagothai Research Centre, University of Pecs, Pecs, Hungary.

Purpose: Investigation of the role of triglyceride metabolism in various diseases, such as cardiovascular or cerebrovascular diseases is still in the focus of numerous studies. Genome-wide association studies reported a number of polymorphisms associated with plasma lipid level changes. The aim of our study was to investigate the distribution of eight variants: rs12130333 at the ANGPTL3, rs16996148 at the CILP2, rs17321515 at the APOE, rs12130332 at the GALNT2, rs4846914 at the FTLD, rs1799941 at the SHBG, rs12130334 at the UBE2D3, rs17999541 at the LPA. After the investigation of polymorphisms in the healthy Roma population samples in the allele frequencies of GALNT2 variant (G allele frequency of rs4846914: 46.6% in Romas vs. 54.5% in Hungarians, p<0.05), while no differences in GALNT2 were identified in healthy Hungarian population samples, however no association with triglyceride levels could be established.

2132F
Arterial Tortuosity in Patients with Vascular Abnormalities due to Mutation in Filamin A. E. Reinstein1, S. Morris2, D. Rimon2, R. Lacro1,2. 1) Medical Genetics Inst, Cedars Sinai Medical Center, LA, CA, USA; 2) Texas Children's Hospital / Baylor College of Medicine TX, USA; 3) Boston Children's Hospital, MA, USA.

Introduction: Arterial tortuosity of the head and neck vessels, as measured by the Vertebral Artery Tortuosity Index (VTI), was recently demonstrated to be elevated in patients with Loeys-Dietz and Marfan syndromes, and correlated with adverse cardiovascular outcomes in those patients. Mutations in the Filamin A (FLNA) locus give rise to several disorders including the otopalatodigital syndrome (OPD; a bone dysplasia), X-Linked Periventricular Heterotopia (XL-PH; a disorder of abnormal neuronal migration), and Cardiac Valvular Dystrophy (CVD). Previous reports also linked mutations in FLNA to a connective tissue disorder characterized by joint, skin and vascular abnormalities, primarily aneurysms of the thoracic aorta. Methods: In this study, we evaluated a mother-daughter pair because of vascular aneurysms due to mutation in the FLNA gene. Calculation of arterial tortuosity following magnetic resonance angiography analysis was performed. Results and Discussion: Significant expansion of aortic aneurysm was seen in the mother despite normal VTI, culminating in an early aortic replacement surgery. In the daughter, a slight expansion in the size of a subclavian artery aneurysm has been observed and surgery is pending. These findings suggest that vascular surveillance is warranted in patients with arterial aneurysms due to FLNA mutations despite the presence of normal vascular tortuosity.

2133W
Phenotype, Genotype and Natural History of Arrhythmogenic Dilated Cardiomyopathy. A. Spezzacatene1,2, G. Sinagra3, M. Merlo4, G. Barbati2, D. Slavov5, A. Di Lenarda5, X. Zhu6, S. Graw1, E. Salcedo1, M. Taylor1, L. Mestroni1. 1) Cardiovascular Institute, University of Colorado, Colorado; 2) Cardiovascular Department, Hospital and University of Trieste, Italy.

BACKGROUND Arrhythmogenic dilated cardiomyopathy (aDCM) is a form of dilated cardiomyopathy (DCM) frequently encountered in clinical practice and characterized by severe venricular arrhythmias. The purpose of this study was to determine the prevalence, genotype-phenotype correlations and long-term outcome predictors of aDCM, to implement current criteria for risk-stratification. METHODS From February 1979 to November 2012, we studied 461 patients (364 families) with DCM enrolled in the Familial Cardiomyopathy Registry. Criteria for aDCM were DCM with VT, SD, AICD shock, syncope, cardiac arrest, PVC>1000/24h. A subgroup of patients was tested for DCM genes (MYH6, MYH7, TNNT2, TTN, LMNA, LAP2, SCN5A, DES). For survival analysis, endpoints were: 1) death or heart transplant, 2) death, heart transplant or malignant ventricular arrhythmias (MVA), 3) MVA. RESULTS Over a follow-up of 96±82 months (up to 20 years), we identified 211 patients (45.8%) with aDCM: among them 72 (15.6% of the total DCM population) experienced MVA during follow-up. Compared to DCM patients, aDCM patients had worse survival (endpoint 1: P=0.006; endpoint 2: P<0.001). Independent risk factors for endpoint 1 in the aDCM population were LVEF<34% and the presence of QRS>110 ms. For survival analysis, 21 patients (4.5%) with aDCM among them 72 (15.6% of the total DCM population) experienced MVA during follow-up. Compared to DCM patients, aDCM patients had worse survival (endpoint 1: P=0.006; endpoint 2: P<0.001). Independent risk factors for endpoint 1 in the aDCM population were LVEF<34% and the presence of QRS>110 ms in V1-V3. There was a cumulative risk for death or heart transplant per additional risk factor. Predictors of MVA in the aDCM population were the presence of QRS>110 ms in V1-V3 and family history of MVA, aDCM was most frequent among LMNA (70.6%), TTN (54.5%) and SCN5A variant carriers (50%). Furthermore, nuclear envelope (LMNA and LAP2) variant carriers had worse NYHA (P=0.030) in spite of smaller LVEDD at baseline (P=0.003), and experienced more heart transplants (P=0.010). TTN carriers had worse event-free survival for endpoint 2 (P=0.037). Finally, SCN5A mutation carriers were younger at enrollment (P=0.004) compared to non-carriers, had prevalence of males (P=0.029) and epsilon waves (P=0.030).

CONCLUSIONS In a large and extensively studied DCM cohort, we define a novel subpopulation characterized by prominent ventricular arrhythmia and different prognosis. Our results suggest that aDCM may benefit from more aggressive therapeutic interventions including modified ICD criteria.
Advancing Genomic Research on Congenital Cardiac Malformations in Africa. P. Kruszka, M. Muenke, B. Solomon, P. Lwabi, L. Harris, A. Beatón, C. Sable, B. Langner. 1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Uganda Heart Institute, Kampala, Uganda; 3) Department of Genetics and Metabolism, Children’s National Medical Center, Washington, D.C.; 4) Department of Cardiology, Children’s National Medical Center, Washington, D.C.

Background: Congenital heart disease (CHD) is the most common birth defect and is the leading cause of mortality among congenital malformations. Many types of CHD are associated with syndromes; some isolated genes have been identified in non-African populations. Previously, we have identified mutations in several CHD-related genes: CFC1, GDF1, and FOXH1. However, many causes of CHD remain unknown. Our study will harness emerging genomic technologies to better understand the etiology of CHD in African and other populations. Also, genotype-phenotype analyses will characterize syndromes with known genetic etiologies that have yet to be characterized in African populations.

Methods: Using comprehensive medical genetic/family history, dysmorphology exam, and echocardiogram, CHD patients were phenotyped and assigned clinical diagnoses when possible. In the future, we will conduct trio-based high-throughput genomic investigation (including exome/genome sequencing and copy number analyses). Results: From 2012-2013 we clinically evaluated 211 Ugandan patients with CHD. Forty-eight (22.7%) were clinically diagnosed with a known syndrome. Of these 48, 20 (41.7%) were suspected to have 22q11 deletion syndrome, 15 (31.3%) had convincing signs of aneuploidy, and 9 (18.8%) were diagnosed with Noonan syndrome or related RASopathies. Other syndromes recognized included Williams, Holt-Oram, and Kabuki. Notably, truncus arteriosus was four times more frequent than expected from known prevalence in other populations. Conclusions: This study is anticipated to advance our understanding of CHD, both by characterizing known syndromes in this understudied population as well as revealing novel disease loci in the African cohort. Through this international collaboration, we will pilate the practice of clinically-oriented genomic medicine in a resource-limited setting, the ultimate goal of which is to assist local clinicians and researchers in their management of these complex cohorts of patients.
2138F

Sequence variants in the mitochondrial D-loop region are associated with ventricular arrhythmias and appropriate ICD-therapy. H. Tao, S. Rosenberg, J. wingrove, CardixoDx, Palo Alto, CA94043, CA.

Background: Sudden death from cardiac ventricular arrhythmias is a leading cause of mortality in the industrialized world. Mitochondrial dysfunction has been implicated in both heart failure and arrhythmias. Mitochondrial sequence analysis was evaluated in a primary prevention heart failure (HF) population with implantable cardioverter-defibrillators (ICDs).

Materials and Methods: Mitochondria DNA (mtDNA) was analyzed in subjects enrolled in DISCERN, a multi-center clinical trial (www.clinicaltrials.gov; NCT00507078) designed to identify markers associated with ventricular tachycardia or fibrillation (VT or VF). Cases were defined as subjects with adjudicated ICD therapy: controls were subjects with no documented ICD therapy for ≥2 years. MiDNA was initially sequenced in pooled case/control samples (Set 1; 19 subjects/pool) using the Ion Torrent PGM; the mitochondrial D-loop region was subsequently re-sequenced in individual Set 1 subjects plus an additional 40 subjects (Set 2). Melting temperatures (Tm) in candidate regions was assessed by SYBR Green RT-PCR across Sets 1, 2 and a larger set of 260 subjects (Set 3).

Results: A region within the D-loop, spanning 160-280 bp of the mitochondrial genome, showed significant sequence variant enrichment in the control pool (p=0.007). The Tm of this region was significantly elevated in individual Set 1 control samples (p=0.016); this increase was validated in two additional independent sets of subjects (Set 2: n=40, p=0.007; Set 3: n=260; p=0.0025). Sequencing of the entire D-loop region in individual Set 1 and 2 subjects showed increased sequence variant enrichment in control samples (n=78, p=0.0025); this increase was also significantly associated with lower Tm (ΔTm = 5°C) in the D-loop region.

Conclusion: The number of variants in the mitochondrial D-loop is associated with ventricular arrhythmias and appropriate ICD therapy, and may affect D-loop formation and mitochondrial function.

2139W


Background: Several genetic models for predicting coronary artery disease (CAD) risk have recently been published. We previously validated a genetic model for diagnosis of obstructive coronary artery disease (CAD) risk recently been published. We previously validated a genetic model for diagnosis of obstructive coronary artery disease (CAD) risk in patients with obstructive CAD by AUC analysis. To clinical factors and gene expression significantly increased the ability to detect obstructive CAD by AUC analysis.

Materials and Methods: A region within the D-loop, spanning 160-280 bp of the mitochondrial genome, showed significant sequence variant enrichment in the control pool (p=0.007). The Tm of this region was significantly elevated in individual Set 1 control samples (p=0.016); this increase was validated in two additional independent sets of subjects (Set 2: n=40, p=0.007; Set 3: n=260; p=0.0025). Sequencing of the entire D-loop region in individual Set 1 and 2 subjects showed increased sequence variant enrichment in control samples (n=78, p=0.0025); this increase was also significantly associated with lower Tm (ΔTm = 5°C) in the D-loop region.

Conclusion: The number of variants in the mitochondrial D-loop is associated with ventricular arrhythmias and appropriate ICD therapy, and may affect D-loop formation and mitochondrial function.

2140T


Introduction: Interpretation of genetic screening results in Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy (ARVD/C) is often difficult. Pathogenicity of variants with uncertain significance may be predicted by software algorithms. However, functional assessment can unambiguously demonstrate the effect of such variants.

Aim: Functional analysis of putative splice site variants identified in ARVD/C patients. Methods: Eleven variants in desmosomal (PKP2, JUP, DSG2, DSC2) and non-desmosomal (TMEM43) genes with potential RNA splicing effect (predicted by SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, and Human Splicing Finder) were analyzed. The variants were found in patients with ARVD/C (fulfillment of 2010 Task Force Criteria) or suspected ARVD/C (3 patients: DSC2 c.1350A>G, JUP, TMEM43 variant). Total RNA was isolated from fresh blood samples and subjected to nPCR. Obtained cDNA products were amplified by RT-PCR with specific primer set. All cDNA fragments were separated according to size using gel electrophoresis. PCR fragments were subjected to direct sequence analysis. Results: Of the 11 variants, 6 were intron and 5 exonic. Eight variants, including 2 missense variants, had a functionally deleterious effect on splicing, heterogeneity present in mRNA splicing by causing exon skipping, generating new splice sites, or activating cryptic sites. All 6 intronic variants tested affected mRNA splicing. Two of 5 exonic variants severely impaired pre-mRNA processing. The remaining 3 exonic variants had a detectable effect on splicing, heterogeneity present in mRNA confirmed biallelic expression in these cases.

Conclusion: Eight variants of uncertain significance had a functional effect on mRNA splicing, indicative of being ARVD/C related pathogenic splice site mutations. This highlights the importance of functional assessment of putative splice site variants to enhance patient care and facilitate cascade screening.

2141F

Functional analysis of collagen VI variants and their contribution to atrioventricular septal defect in Down syndrome. A. Ackerman1, P. Holdenh1, J. Fitzgerald2, C. Maslen1, 1) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR; 2) Department of Orthopedics & Rehabilitation, School of Medicine, Oregon Health & Science University, Portland, OR.

Atrioventricular septal defect (AVSD) is a common heart defect frequently associated with Down syndrome (DS). Children with DS have a 2000-fold increased risk of the defect, yet many have a normal heart indicating that there is another etiology. The focus of our work is on the collagen VI (COL6A1 and COL6A2) genes. Results from our previous reports in DS and non-DS populations have both highlighted the importance of functional assessment of potential splice site mutations. All 11 variants, 6 were intron and 5 exonic. Eight variants, including 2 missense variants, had a functionally deleterious effect on splicing, heterogeneity present in mRNA splicing by causing exon skipping, generating new splice sites, or activating cryptic sites. All 6 intronic variants tested affected mRNA splicing. Two of 5 exonic variants severely impaired pre-mRNA processing. The remaining 3 exonic variants had a detectable effect on splicing, heterogeneity present in mRNA confirmed biallelic expression in these cases. We hypothesized that AVSD-associated variants in COL6A1 and COL6A2 impact collagen VI biosynthesis disrupting microfibril formation in the extracellular matrix in the developing heart that leads to AVSD. To determine if rare variants play a pathogenic role in AVSD we analyzed the biosynthesis and trafficking of proteins expressed by three missense variants in COL6A1 (V117A, Q768H, and R872W) and two variants in COL6A2 (p.E106K and p.R853Q). Immunoblot analysis showed that the α1 protein expressed by the COL6A1 variant p.R872W was undetectable in the media, and α2 proteins for both COL6A2 variants, p.E106K and p.R853Q, were observed in reduced amounts in the media compared to wildtype (WT). In addition, the α2 protein expressed by the R853Q variant was retained in the cell, compared to WT. To test if the α1 chain produced by missense variants were triggering cellular processes activated by endoplasmic reticulum (ER) stress, we analyzed expression of HERPUD, which directs misfolded proteins to the proteosome for degradation, and CHOP, a pro-apoptotic transcription factor activated by ER stress. We found that the cells expressing p.R872W had increased expression of V117A, p.Q768H, and p.R872W variants expressed significantly increased levels of HERPUD than WT α1. Cells expressing p.Q768H and p.R872W variants also expressed significantly higher levels of CHOP compared to WT. Our results indicate that rare variants in COL6A1 and COL6A2 associated with AVSD cause defective protein secretion and intracellular retention of protein chains that form collagen VI microfibrils. In addition, these variants are coupled with increased expression of proteosomal degradation and pro-apoptotic genes as a result of ER stress. Further analysis of these variants will provide mechanistic insight into how collagen VI contributes to heart defects during development.
Epigenomic regulation in the promoter of miR-210 gene affects HIFα binding and vascular diseases. S.H. Joo1, K.C. Chen1, Y.S. Wang1, Y.C. Liao2-3, 1) Genome Medicine, Kaohsiung Medical Univ, Kaohsiung City, Taiwan; 2) Taichung Veterans General Hospital, Taichung, Taiwan; 3) Department of Neurology, National Yang-Ming University School of Medicine, Taipei, Taiwan.

microRNAs are short non-coding RNAs that can regulate gene expression by inhibiting target mRNAs. Therefore, microRNAs are considered to play a regulatory role for gene expression. However, the regulation of microRNA expression is mainly unknown. In this study, we presented the data to indicate DNA methylation can control microRNA expression. microRNA-210 (miR-210) has been reported to be involved in hypoxia-related events. Several studies suggested that miR-210 expression can be regulated by hypoxia-inducible transcription factors-1αα (HIF-1αα) during hypoxia. Further, by using the bioinformatics analysis, the promoter of the miR-210 gene contains several CpG-rich regions. However, there are no reports about epigenetic regulation of miR-210 gene. We investigated the role and mechanism of miR-210 in the cardiovascular system. We first found miR-210 gene was upregulated in vascular smooth muscle cells (VSMC) and human umbilical venous endothelial cells (HUVEC) when they were exposed to oxidized low-density lipoprotein (oxLDL). By combining with 5-Aza-2’-deoxycytidine treatment or DNA methyltransferase 3b knockdown, we found that miR-210 gene expression could be epigenetically regulated by oxLDL. By using the bisulfite sequencing assay, the decreased methylation levels of miR-210 gene was identified in vitro and in vivo. This methylation change could affect the HIF-1αα-regulated miR-210 gene expression. Furthermore, over-expression or knockdown of miR-210 influenced the activity of oxLDL-VSMCs and HUVECs migration. According to the bioinformatic prediction, SPRED2 (an inhibition regulator in ERK activation) could be a major miR-210’s target gene involved in cell migration. Our reporter assays using wild-type and mutant 3’ UTR of SPRED2 gene confirmed that SPRED2 is a novel direct target of miR-210. Over-expression or knockdown of SPRED2 affected oxLDL-mediated ERK/c-Fos/MMPs pathways in cell migration. Altogether, the data suggest that epigenetic regulation of miR-210 gene and its effect on cell migration are involved in the cardiovascular diseases formation.


Hepatic lipase is an enzyme involved in the metabolism and regulation of the plasma lipoproteins. Variants in the HL gene (LIPC) may influence plasma lipoprotein levels. In this study, distribution of LIPC C-514T and G-250A polymorphisms and their associations with plasma lipid profile in young Jordanians was investigated. Genotyping of the polymorphisms in 348 young Jordanian adults was performed using polymerase chain reaction-restriction fragment length polymorphism technique. The G-250 and C-514 alleles are abundant in the Jordanian population with frequencies of 79% and 80%, respectively. However, no significant difference in lipid-lipoprotein profile (total cholesterol, triglycerides, LDL and HDL) between the different genotype groups of either C-514T or G-250A polymorphisms even when males and females were examined separately (P > 0.05). Thus, in young Jordanian adults, the examined LIPC polymorphisms seem to play a limited role in determination of plasma lipoprotein levels.

Novel mutation in PRKAG2 gene highlights the allosteric site of AMPK. C. Phromphutkul1, A. Gray2, K. Rotondo2, J. Pambud2, Y. Tseng3, D.G. Hardie2,1) Pediatrics, Rhode Island Hosp, Providence, RI; 2) College of Life Sciences, University of Dundee, Dundee, Scotland, UK; 3) Pediatrics, Women and Infants’ Hospital, Providence, RI.

Mutations in PRKAG2 have been implicated in hypertrophic cardiomyopathy (HCM). PRKAG2 encodes a γ2 subunit of the nucleotide-binding regulatory subunit of AMP-activated protein kinase (AMPK), a heterotrimeric enzyme with major roles in regulation of energy metabolism in response to cellular stress. Twelve different heterozygous point mutations within the nucleotide-binding domains have been reported. De novo mutations (R531Q and R384T) were associated with severe disease and death during infancy. Case study: An abnormal 27-week prenatal ultrasound, consistent with HCM, was noted. At birth, the cardiac echo confirmed HCM. Parents had normal cardiac evaluation. Molecular testing for HCM panel was conducted and a de novo mutation in the PRKAG2 gene [Lys475Glu (K475E) likely disease causing] was identified in the child. At three years of age, the child has modest HCM but clinically is doing well on supportive medications. Unlike the previous cases with R531Q and R384T mutations, this case appears to have a better clinical course. To investigate the significance of the K475E mutation on AMPK complex, functional studies were performed. Method: HEK-293 stably expressing wild type (WT) or K475E FLAG-tagged γ2 from a tetracycline-dependent expression system were established. AMPK complexes containing the WT or K475E mutant γ2 were immunoprecipitated using anti-FLAG antibodies and their activities determined at varying AMP concentrations. We also compared the effects of metabolic stress induced using phenformin. Results: Our molecular studies show that the K475E mutation has three effects on the regulation of the AMPK complex: (i) markedly increasing the basal phosphorylation of Thr-172 and associated kinase activity; (ii) reducing the sensitivity to AMP in allosteric activation; (iii) preventing the increased Thr-172 phosphorylation and activity observed in response to the metabolic stressor, phenformin, in intact cells. Conclusions: K475E is conserved in γ1 and γ2 isoforms of all species, and examination of crystal structures shows that its side chain could form electrostatic interactions with the phosphate groups of AMP, ADP or ATP bound in site 1. The mutation almost completely abolishes allosteric effects of AMP, ADP or ATP on K475E and Thr-172 phosphorylation and activity and prevents further activation by metabolic stress. Ongoing studies of the effect of these changes on downstream pathways are underway in patient fibroblasts.

Systematic cell-based functional screening for novel cardiovascular risk genes. H. Runz1,2,4, C. Schubert1,3, P. Blattmann2-3, G. Domschke2, A. Thormaehlen1-3, K.C. Chen1,3,4, K. Rotondo1,2,3, Y. Tseng3,1) Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany; 2) European Molecular Biological Laboratories (EMBL), Heidelberg, Germany; 3) Molecular Medicine Partnership Unit (MMPU), Heidelberg, Germany; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston, USA.

Several studies have highlighted the potential role of genetic factors in the development of cardiovascular diseases. Large population cohorts are uncovering numerous genes and genetic variants with putative relevance for blood LDL-C levels and predisposition to disease. However, discovery of novel disease genes and improvement of therapies is challenged by the fact that (1) most candidate genes harbor vast numbers of potentially relevant, yet rare and often neutral variants that reduce the power of association testing; and (2) a modest mechanistic understanding of how genes and variants impact relevant phenotypes. To address these limitations, we have established technology that complements human genetics by systematic, quantitative phenotypic analyses of gene and variant function in cultured human cells. Our unbiased, scalable functional genomics strategy is based on reducing (by RNAi-interference), increasing (by cDNA-overexpression) or complementing (by cDNA-complementation) the cellular levels of candidate genes, high-throughput functional assays, automated microscopy and multiparametric image analysis. We will present data on how we are applying this technology to systematically characterize >100 candidate genes suggested as likely associated with blood lipid levels, coronary artery disease and/or myocardial infarction for functions in regulating LDL-C uptake and turnover in cells. We will provide mechanistic insight how several genes without previously known lipid-regulatory role may affect levels and function of the LDL-receptor and provide hypotheses how variation in these genes could predispose to disease. Our study proposes that the combination of genetics with systematically-acquired functional data in cells can pinpoint new genes responsible for altered LDL-C and MI-risk.
Using Genetic Information to Analyse and Define the Complex Aetiology of Blood Pressure. K.H. Wade, N.J. Timpson, G. Davey Smith. School of Social and Community Medicine, University of Bristol, Bristol, Bristol, United Kingdom.

Blood pressure (BP) is an important phenotype that has had its study hampered by complexity. The discovery of 29 SNPs, robustly associated with BP, has provided an opportunity to examine the aetiology of BP through SNP-based associations to further understand the mechanisms that maintain and alter BP at different stages of development. We aimed to assess the contribution of 29 SNPs (individually and in a weighted score) reported in the International Blood Pressure Consortium GWAS on BP in children (mean age 3-18 years). The Avon Longitudinal Study of Parents and their Children is a geographically-based birth cohort investigating factors influencing health and development of children. After data cleaning, 8,361 unrelated individuals had genome-wide information and at least one systolic and diastolic BP (SBP and DBP, respectively) measured at 11 times (mean ages: 3-5, 8, 10-15, and 18). We examined the associations between SNPs and BP at different ages in the same samples cross-sectionally and longitudinally, and assessed whether the overall contribution of SNPs changed with age. Cross-sectionally, the score was associated with SBP at ages 8-15 years and with DBP at ages 8-18 years. For example, the score was associated with a 12.67 mmHg increase (95% CI: 4.88, 20.46; p = 0.001) in SBP at age 11 and an 18.87 mmHg (95% CI: 8.18, 29.56; p = 0.001) increase in SBP at age 15. Individually, few SNPs showed associations with BP cross-sectionally. At age 11, rs7129220 was associated with a 0.68 mmHg (95% CI: 0.12, 1.25; p = 0.018) increase in SBP and rs1799945 was associated with a -0.51 mmHg change (95% CI: -0.99, -0.04; p = 0.035). Longitudinally, age was strongly associated with a 0.0377 mmHg (95% CI: 0.0372, 0.0382; p < 0.0001) increase in SBP and a 0.0138 mmHg (95% CI: 0.0134, 0.0142; p < 0.0001) increase in DBP, with no evidence of interaction between score (separated into centiles) and age (p = 0.890 for SBP, p = 0.223 for DBP). However, there was evidence of interaction between 6 SNPs and age (p = 0.031 for rs143076; 0.048 for rs805303 and DBP; 0.011/0.04 for rs932764 and SBP/DBP, respectively; 0.036 for rs1327235 and SBP; 0.016 for rs3774372 and SBP; 0.032 for rs17249754 and SBP, and 0.02 for rs12940887 and SBP). We have further investigated the interactions between 6 SNPs and age in the first 11 years of the study and the results will be presented to the audience.
2148W

Analysis of vWA and TPOX loci reveals a highly associated relationship with thrombosis disease in Mexican Mestizo population. R. Camacho1, A. Majluf2, A. Maguię2, G. Noris3, C. Santana3, M.A. Moraz3, J. Hernández1, R. Gómez1. 1) Departamento de Toxicode, Cinvestav-IPN, México D.F.; 2) Unidad de Investigación Médica en Trombosis y Aterogénesis, Instituto Mexicano del Seguro Social, México D.F.; 3) Laboratorio BIMODI (Biología Molecular Diagnóstica), Querétaro, Qro; 4) Departamento de Biomedicina Molecular, Cinvestav-IPN, México D.F.

Thrombosis disease is an important target for biomedical research due to its high morbidity and mortality around the world. Thrombosis is a complex disease and several risk factors are evolved (age, trauma, hormonal alterations, immobility, obesity, diabetes type 2, and hypertension) as well as environmental factors. In addition, genetic factors are strongly associated with this disease. Recent studies suggest that ethnicity is considered an important risk factor in thromboembolism incidence, showing a variable frequency of the disease among different ancestors. In this study we identify truly genetic biomarkers, type short tandem repeat (STRs), associated with thrombosis in a population with multiethnic backgrounds from Mexico. Genotypes were obtained using a validated method of genotyping from patients (n=177) with venous thrombosis. The allelic frequencies were compared with non-related healthy individuals (n=531). Bayesian methods were used to correct population stratification in order to avoid spurious associations due to admixture. We found that vWA-1b (OR=1.5), TPOX-9 (OR=1.7), and TPOX-12 (OR=2.3) alleles were significantly associated with thrombosis disease, showing accurate confident intervals. This association remains after admixture correction, confirming a strong genetic association (P<0.0001). In addition, the combination of allele 16 (vWA) with allele 12 (TPOX), showed a strong effect when compared with healthy individuals (1.02 - 3.64), suggesting that this combination could be associated with an increase of two times of the thrombosis risk. In summary, our data propose that vWA and TPOX are good biomarkers, which may improve in the diagnostic methods significantly treatment costs by providing preventive and personalized medicine.

2149T

GALNT2, a HDL cholesterol and triglyceride candidate gene, regulates lipoprotein metabolism in vivo. S. Khetarpal1, A. Edmondson1, A. Raghavan3, S. Khatiresan2, D. Rader1. 1) Dept Med, Univ Pennsylvania, Philadelphia, PA; 2) Broad Institute and Massachusetts General Hospital, Boston, MA.

Several genome-wide association studies (GWAS) for blood lipids have uncovered novel genomic loci influencing these traits. One such locus associated with high-density lipoprotein cholesterol (HDL-C) and triglycerides (TG) is that harboring GALNT2 on chromosome 1q42. GALNT2 encodes GalNac-T2, an enzyme catalyzing the initiation of O-glycosylation of specific protein targets. We previously demonstrated an in vivo role for GALNT2 in lipoprotein metabolism through GALNT2 overexpression and knockdown in mice using adenov-associated virus vectors. To further study the mechanism through which GALNT2 regulates plasma lipids, we have generated whole-body and liver-specific GALNT2 knockout (KO) mice. Whole-body GALNT2 deficiency resulted in elevated VLDL-TG levels compared with wild-type and a moderate reduction in HDL-C (20%, P < 0.05). GALNT2 KO mice did not exhibit differences in post-prandial TG clearance or VLDL-TG secretion. Liver-specific deletion of GALNT2 reduced HDL-C (26% decrease relative to WT, P < 0.01), and reduced numbers of HDL particles (19% decrease relative to WT, P < 0.01) with minimal changes in plasma TG or dietary TG clearance. Our data further implicate GALNT2 as an important regulator of plasma lipoprotein metabolism with tissue-specific regulatory roles in vivo.

2150F

Disease-relevant pathways modulate a cis-regulatory element at the TCF21 coronary heart disease locus. C. Miller1, D. Anderson1, R. Kundu1, R. Raesdijk1, S. Nömburg1, R. Diaz1, N. Leeper1, E. Schadt3, C. Hsing2, T. Assimes1, T. Quertermous1. 1) Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 2) Division of Biostatistics and Bioinformatics, National Health Research Institutes, Zhunan, Taiwan; 3) Institute for Genomics and Multiscale Biology, Mount Sinai School of Medicine, New York, NY.

Coronary heart disease (CHD) is the leading cause of mortality in both developed and developing countries worldwide. Genome-wide association studies (GWAS) have now identified 46 independent susceptibility loci for CHD, however the biological and disease-relevant mechanisms for these associations remain elusive. The large-scale meta-analysis of GWAS recently identified in Caucasians a CHD-associated locus at chromosome 6q23.2, a region containing the transcription factor TCF21 gene. TCF21 (Capsulin/ Pod1/Epicardin) is a member of the basic-helix-loop-helix (bHLH) transcription factor family, and regulates cell fate decisions and differentiation in the developing coronary vasculature. Herein, we characterize a cis-regulatory mechanism by which the lead polymorphism rs12190287 disrupts an atypical activator protein 1 (AP-1) element, as demonstrated by allele-specific transcriptional regulation, transcription factor binding, and chromatin organization. Further, this element is shown to mediate signaling through platelet-derived growth factor receptor beta (PDGFRβ) and Wnt’s tumor 1 (WT1) pathways. A second disease allele identified in East Asians also appears to disrupt an AP-1-like element. Thus, both disease-relevant growth factor and embryonic signaling pathways may regulate CHD risk through two independent alleles at TCF21.

2151W

A functional polymorphism that affects the APOA5 gene expression determines plasma triglyceride levels conferring coronary atherosclerosis risk in Han Chinese Population. W. Shou1, F. Xie2, Y. Wang3, B. Wang1, Z. Wang1, W. Huang1, 1) Department of Genetics, Shanghai MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center, Shanghai, China; 2) Rui Jin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China.

Elevated plasma triglyceride (TG) level has been established as an independent risk factor for multiple cardiovascular diseases. Thus, it is clinically important to identify genetic determinants of plasma TG levels. APOA5 in 11q23.3-q23 gene cluster, the top hit of the TG-associated loci, encodes an apolipoprotein APOA5, which promotes lipoprotein lipase mediated plasma TG hydrolysis. A polymorphism rs2266788 in 3’ untranslated region of the APOA5 gene was identified significantly associated TG levels, but was underrepresented compared to the lead variants in genome-wide association studies. So, we investigated the association between rs2266788 and plasma TG regulation, and between rs2266788 and coronary artery disease (CAD) incidence in Han Chinese population. rs2266788 was genotyped in 3222 unrelated subjects consisting of 2,062 CAD cases and 1,160 matched controls in China. In stratified analyses, the association was all significant in controls, in cases of CAD, and in females. In logistic regression analyses adjusted for age, sex and CAD status, Plasma TG levels were higher in samples with C allele than T allele. Elevated plasma triglyceride (TG) level has been established as an independent risk factor for multiple cardiovascular diseases. Thus, it is clinically important to identify genetic determinants of plasma TG levels. APOA5 in 11q23.3-q23 gene cluster, the top hit of the TG-associated loci, encodes an apolipoprotein APOA5, which promotes lipoprotein lipase mediated plasma TG hydrolysis. A polymorphism rs2266788 in 3’ untranslated region of the APOA5 gene was identified significantly associated TG levels, but was underrepresented compared to the lead variants in genome-wide association studies. So, we investigated the association between rs2266788 and plasma TG regulation, and between rs2266788 and coronary artery disease (CAD) incidence in Han Chinese population. rs2266788 was genotyped in 3222 unrelated subjects consisting of 2,062 CAD cases and 1,160 matched controls in China. In stratified analyses, the association was all significant in controls, in cases, in females and in males. In logistic regression analyses adjusted for age, sex and CAD status, Plasma TG levels were higher in samples with C allele than T allele. Elevated plasma triglyceride (TG) level has been established as an independent risk factor for multiple cardiovascular diseases. Thus, it is clinically important to identify genetic determinants of plasma TG levels. APOA5 in 11q23.3-q23 gene cluster, the top hit of the TG-associated loci, encodes an apolipoprotein APOA5, which promotes lipoprotein lipase mediated plasma TG hydrolysis. A polymorphism rs2266788 in 3’ untranslated region of the APOA5 gene was identified significantly associated TG levels, but was underrepresented compared to the lead variants in genome-wide association studies. So, we investigated the association between rs2266788 and plasma TG regulation, and between rs2266788 and coronary artery disease (CAD) incidence in Han Chinese population. rs2266788 was genotyped in 3222 unrelated subjects consisting of 2,062 CAD cases and 1,160 matched controls in China. In stratified analyses, the association was all significant in controls, in cases, in females and in males. In logistic regression analyses adjusted for age, sex and CAD status, Plasma TG levels were higher in samples with C allele than T allele.
2152T
Use of allele-specific FAIRE for identification of functional variants at cardiometabolic loci. A.J.P. Smith1, J. Palmén1, E. Romeo2, F. Drenos1, P. Howard1, A.D. Hingorani2, P.J. Talmud2, S.E. Humphries1. 1) Cardiovascular Genetics, University College London, London, United Kingdom; 2) Genetic Epidemiology, University College London, London, United Kingdom.

GWAS for cardiometabolic diseases have identified many novel loci for further analyses, although identification of functional, non-coding variants at the majority of these loci remains limited. Using FAIRE followed by allele-specific analysis with the Cardiometabochip array, we have examined GWAS loci for effects on chromatin accessibility as a marker of regulatory potential, to differentiate functional variants from non-functional variants in strong LD. Using this methodology, we have localised functional SNPs at several loci. We confirm the functionality of two variants at the 9p21 locus, previously implicated in coronary artery disease and type 2 diabetes susceptibility, and describe their effects on protein-DNA interactions, reporter gene expression and biomarkers.

2153F
Identification and characterization of patients with autosomal dominant hypercholesterolemia caused by gain-of-function mutations in proprotein convertase subtilisin/kexin type 9 and comparison with patients with Familial Hypercholesterolemia (FH) and Familial Defective apolipoprotein B (FDB).

Gain-of-function mutations (GoFm) in proprotein convertase subtilisin/kexin type 9 (PCSK9), a potent modulator of LDLR, and early onset cardiovascular disease. The most frequent mutations causing ADH are found in the LDL receptor (LDLR; FH) or its ligand apolipoprotein B (ApoB; FDB). Gain-of-function mutations (GoFm) in proprotein convertase subtilisin/kexin type 9 (PCSK9), a potent modulator of LDLR on hepatocytes, appear in ~2% of patients with ADH. To date, relatively few patients with PCSK9 GoFm have been described, and their clinical syndrome incompletely characterized. To better understand the geographic and familial distribution of PCSK9 GoFm, their clinical manifestations, and their comparison to FH and FDB we conducted a retrospective, cross-sectional parallel-group observational cohort study. Patients with PCSK9 GoFm confirmed by molecular testing were matched for age and sex with patients with molecularly proven FH and FDB. Data collected included baseline and on-treatment lipid profiles, the presence of xanthoma, xanthelasma, and corneal arcus, and the occurrence and age of onset of CVD. We initiated 200 site contacts and collected data on 164 PCSK9 GoFm patients (83 men, 81 women) from 12 sites in France, Japan, Norway, Portugal, South Africa, The Netherlands, the UK, and the USA. We matched these patients with 2126 patients with FH and 470 with FDB. We characterized the LDLR mutations as “defective” (missense, small in-frame indel, synonymous with added splice site) or “deficient” (large or frame-shifting indel, nonsense, splice site, promoter snp) and compared their lipid profiles. In general, individual PCSK9 GoFm appeared to have restricted geographic distributions and to be concentrated within a small number of pedigrees. Examples include 22 patients with R215H found only in 2 pedigrees in Norway, and 12 patients with V41 and 30 patients with E32K found only in Japan. These data support a similar (recent origin) history for PCSK9 GoFm and LDLR FH mutations that is distinct from ApoB FDB mutations. Mean baseline LDLC was highest in patients with PCSK9 GoFm and lowest in those with FDB. Of the patients with PCSK9 for whom data were available, 53% had evidence of xanthoma, and 33% had a history of CAD. Index patients with either FH or FDB had higher baseline LDL levels than family members and patients with deficient LDLR mutations had higher baseline LDLC than those with defective mutations.

2154W
GATA4 is implicated in the pathogenesis of neonatal and childhood-onset diabetes. E. De Frange1, C. Shaw-Smith1, H. Lango Allen1, S. Flanagan1, M. Borowiec1, W. Mylnarski2, M. Ballo1, J. Ferretti3, A. Hattersley1, S. Ellard1. 1) University of Exeter Medical School, Exeter, United Kingdom; 2) Department of Paediatrics, Oncology, Haematology and Diabetology, Medical University of Lodz, Lodz 91-738, Poland; 3) Genomic Programming and Analysis-Cells Laboratory, Institut d’Investigacions August Pi i Sunyer (IDIBAPS), Barcelona, Spain.

Recent studies have shown that GATA6 haploinsufficiency is the most common cause of pancreatic agenesis and can also result in diabetes diagnosed outside the neonatal period. GATA4 is a transcription factor closely related to GATA6 and is known to be involved in the first stages of pancreatic development in mouse models: conditional knock-out of Gata4 in mouse embryos causes agenesis of the ventral pancreas. In humans GATA4 has been long regarded as a candidate regulator of pancreatic development but the evidence reported so far has been inconclusive. A single patient with a congenital heart defect and pancreatic agenesis harboring a GATA4 missense mutation has been reported, but causality was not proven. Similarly to GATA6, GATA4 mutations and deletions are a known cause of congenital heart defects.

We report four patients with diabetes, congenital heart malformations, and deletions including the GATA4 gene locus. In three cases diabetes presented in the neonatal period (age at diagnosis: 1-7 days) and the fourth case was diagnosed at 13 years. In addition we identified a fifth patient with a de novo GATA4 missense mutation (p.Asn273Lys) with complete absence of the pancreas confirmed post mortem. This mutation affects a highly conserved residue located in the second zinc finger domain of the GATA4 protein. Functional in vitro studies showed reduced DNA binding and transactivational activity of the mutant protein.

These results suggest that GATA4 haploinsufficiency is a cause of neonatal or childhood-onset diabetes with or without pancreatic exocrine insufficiency. GATA4 mutation testing should therefore be considered in patients with neonatal/childhood-onset diabetes and congenital heart defects in whom no GATA6 mutation has been identified.

2155T
SMAD3 haploinsufficiency has a causative role in development of TAAD. L. Gong1, X. Duan1, P. Yang2, X.F. Wang2, O.A. Moffit1, D.M. Milewicz1. 1) Department of Internal Medicine, University of Texas Health Science Center at Houston, 6431 Fannin, MSB 6.100, Houston, TX 77030, USA; 2) Department of Pharmacology and Cancer Biology, Duke University, Durham, North Carolina, USA.

Thoracic aortic aneurysms leading to acute aortic dissections (TAAD) can be inherited in families in an autosomal dominant manner with reduced penetrance and variable clinical expression. Mutations in genes of the TGF-beta signaling pathway are believed as the causes of the disease in these families. The majority of these mutations are predicted to decrease TGF-beta signaling but there is evidence of increased signaling based on phos- phorylation Smad2 (pSmad2) in aortic tissue of patients with these mutations. SMCs were explanted from a patient with SMAD3 R279K and SMAD3-/- mice to investigate TGF-beta signaling. After exposure to TGFbeta1, the patient’s SMCs showed decreased phosphorylation of Smad3 but increased pSmad2 compared with wild-type (WT) SMCs and the Smad3/-/- SMCs also had increased and sustained pSmad2 signaling with exposure to TGF-beta1. When treated with SB-431542, a small molecule inhibitor of the type I TGF-beta receptor, the pSmad2 signaling on Smad3-/- SMCs was reduced. Both Smad3-/- and SMAD3R279K SMCs also had reduced expression of SMC contractile proteins compared with WT SMCs, along with decreased expres- sion of other genes known to be dependent on Smad3 signaling, including Col1A1 and Col3A1. The expression of DCN in the patient’s SMCs was increased. Based on these results it was determined that the mutant Smad3 could not complex with Smad4 but smad3 WT could. In both the transfected and SMAD3 R279K SMCs, SMAD3 remained in the cytoplasm 1 hour after exposure to TGFbeta1, but when WT transfected or WT SMC, SMAD 3 was in the nucleus. The increased tgbf3, which increases tgbf1 transcription, was upregulated in response to the increased smad2 signaling. Therefore, SMAD3 mutations disrupt Smad3 signaling in SMCs thus disrupting expression of contractile proteins and collagen. At the same time, SMCs have increased signaling through smad2, including increased pSmad2 and Smad2/3 phosphorylation. Further studies will determine if the aortic disease results from loss of Smad3 signaling or increased smad2 signaling.
Background: We previously showed that Tnfaip8 was associated with susceptibility to S. aureus infection in susceptible (A/J) but not resistant (C57BL/6J) mice (Ahn, PLOS Pathogens 2010). However, the role of Tnfaip8 in host response to S. aureus is unknown. Methods: RNA and protein expression profiles of both variants of Tnfaip8 in both S. aureus-susceptible (A/J) and resistant (C57BL/6J) mice were evaluated by real-time PCR and western blot. Expression profiles of Tnfaip8 in patients with bloodstream infection (BSI) due to S. aureus (n=32) or E. coli (n=19) or healthy subjects (n=44) were evaluated with whole genome expression data. Results: Levels of Tnfaip8 RNA were respectively 34% (variant 1) and 13% (variant 2) lower in susceptible vs. resistant mice blood (n=6). By western blot lower levels of Tnfaip8 protein was confirmed in A/J mice, but the protein signal was predominated in immune-related tissues (e.g., kidney, spleen, lymph node) in both mouse strains. Both variants of Tnfaip8 were highly expressed in both susceptible and resistant mice blood. However, when challenged with S. aureus or E. coli, whole blood expression of Tnfaip8 variant 1 decreased to non-detectable level in susceptible mice but had more than 100% increase in resistant mice (n=6). Expression of Tnfaip8 variant 2 decreased about 15% in susceptible A/J mice blood, but increased more than 11% in resistant C57BL/6J mice (n=6). Similarly, susceptible patients with S. aureus BSI had 15% decrease (p=0.001) of expression, while E. coli BSI had 12% (p<0.001) decrease of Tnfaip8 as compared to healthy human subjects.

Conclusions: These findings further support the importance of Tnfaip8 in immune response to S. aureus, and suggest that both variants of Tnfaip8 may participate in this function.

2159F


INTRODUCTION Depressive symptoms are associated with increased inflammatory protein levels but only in certain individuals. In a prospective study of patients with acute coronary syndrome (ACS), we tested a biobehavioral model in which inflammatory protein gene polymorphisms interact with depression, and we predicted greater increases in inflammatory protein levels will be observed compared to the levels caused by either gene polymorphisms or depression alone. The purpose of the proposal was to determine a well-defined, high-risk subgroup of ACS patients in which the interaction between depression and genotype polymorphisms increases the risk of subsequent major adverse coronary events (myocardial infarction, revascularization procedures, stroke, and death) more than either factor alone, in part because of their combined effect on increasing inflammatory protein levels. We hypothesized that gene x depression interaction effect on inflammatory protein level will also have a depression-by-gene interaction effect on risk of subsequent major adverse coronary events. METHODS We enrolled ACS patients from two large tertiary care centers and obtained blood samples to measure inflammatory protein levels immediately after hospital admission. Inflammatory proteins and genes measured included Interleukin (IL) 6, C-reactive Protein (CRP), Tumor Necrosis Factor Alpha (TNFα), E-Selectin (SELE), and Monocyte Chemoattractant Protein 1 (MCP-1). Patients were screened for depression using Beck Depression Inventory-II scores. Demographic and clinical risk factors were collected. The hypothesis will be tested using logistic regression and survival analysis. RESULTS Study results are pending data analysis and are currently underway. CONCLUSIONS We will present study findings and discuss how the discovery of a relationship among depressive symptoms, genetics, and inflammatory protein levels in a subgroup of ACS patients provides a rationale for studying environmental triggers of depressive symptoms and the effects of depression interventions such as different medications, psychotherapy, treatment combinations, and self-management techniques like exercise on inflammatory protein levels. The results may lead to specific secondary intervention methods that consider the integrated response of the patient to inflammatory triggers and provide support for the role of genes in modifying the relationship between depression and adverse coronary events.
A genome-wide interaction study identifies PDE1C genetic variants and renal function as jointly associated with coronary artery disease. C. Ward-Caviness 1, M. Winn 1,2, C. Blach 1, C. Haynes 1, E. Dowdy 1, S. Gregory 1, S. Shah 1,2, W. Kraus 1,2, E. Hauser 1, 1) Center for Human Genetics, Duke Univ Med Ctr, Durham, NC; 2) Department of Medicine, Duke Univ Med Ctr, Durham, NC.

Cardiovascular disease (CVD) and chronic kidney disease (CKD) are significant causes of mortality in developed nations. CAD and CKD often co-occur, and each has a well-established genetic basis. However potential interactions between genetic variants (SNPs) and markers of CKD that may influence cardiovascular outcomes have not been explored. We used CATHGEN, a large cardiac catheterization cohort, to perform a genome-wide interaction study that estimated associations between cardiovascular disease, indexed by the number of diseased coronary vessels, and SNP by estimated glomerular filtration rate (GFR) interactions. We used race-stratified cohorts of 2202 whites (EA) and 663 blacks (AA) followed by meta-analysis. A total of 905,956 SNPs were analyzed using a cumulative link model with a logit link. Thus we modeled the log-odds for a change in the number of diseased coronary vessels (our dependent variable/outcome), given a particular SNP-GFR interaction (our independent variable of interest). We adjusted for age, sex, BMI, hypertension, hyperlipidemia, diabetes, smoking, and race-specific principal components to adjust for potential ethnic stratification within each race-stratified cohort. We identified the 10 strongest SNP-GFR interactions for each cohort and looked for those that validated (P<0.05 and consistent direction of association between cohorts), or were among the strongest meta-analysis results, ranked by p-value. Using these criteria we identified rs10951304, an intrinsic PDE1C variant (P=1.09x10−36 AA, P=3.51x10−35 EA, P=9.77x10−36 meta-analysis). Another intrinsic PDE1C variant, rs2058411 (AA P=1.32x10−46), was among the 10 most significant AA associations, however this SNP did not have P<0.05 in the EA. PDE1C is a phosphodiesterase that hydrolyzes cAMP and cGMP, two important cell signaling molecules. PDE1C is responsible for a significant portion of cAMP hydrolytic activity in the heart, particularly in smooth muscle cells cultured from atherosclerotic lesions. These analyses take a novel approach to understanding the co-occurrence of CVD and CKD. Using our genome-wide interaction approach to understanding the co-occurrence of CVD and CKD. Using our genome-wide interaction approach we have identified genetic variants in a portion of cAMP hydrolytic activity in the heart, particularly in smooth muscle important cell signaling molecules.
2162F
Saturated fat intake modulates the association between an obesity genetic risk score and BMI. P. Casas-Agustench 1,2, R. Jiang 3,5, E.R. Parast 3,5, D.K. Nolan 1, B.H. Brunner 1, 2, 3, E.C. de Faria 1, 2

Background and objectives: Combining multiple genetic variants related to obesity into a global genetic risk score (GRS) might improve identification of individuals at risk of developing obesity. Moreover, characterizing gene-diet interactions is a research challenge aimed at investigating dietary recommendations for individuals with higher predisposition to obesity. The aim was to analyze the association between obesity GRS and body mass index (BMI) in a US population, with focus on gene-diet interactions with total and saturated fat intake. Methods: A cross-sectional study including 783 participants from the Genetics of Lipid Lowering Drugs and Diet Network study. A weighted GRS was calculated on the basis of 62 obesity-associated variants. Anthropometrical and biochemical measurements were taken using standard procedures. Dietary intake was estimated with a validated questionnaire. Results: The obesity GRS was evaluated by tertiles for BMI, participants with a higher GRS had a higher BMI (P trend < 0.001). Participants in the highest GRS tertile and with BMI < 30kg/m2 showed lower total fat (2.5%) and SFA (1.0%) intake compared to participants in the same tertile with BMI ≥ 30kg/m2 (P = 0.010 and 0.027, respectively). An interaction was observed between total fat intake and the obesity GRS for BMI (P for interaction = 0.014 for continuously evaluated dietary and GRS variables). We also identified interactions between SFA and MUFA intake and GRS evaluated continuously for BMI and GRS (P = 0.022 and 0.012, respectively). Finally, we obtained similar results when the SFAs were evaluated categorically according to low and high values (P for interaction = 0.014 for categorical SFA intake). Interactions were also observed between total fat, SFA and MUFA intake and the obesity GRS for waist circumference (P for interaction = 0.004, 0.004 and 0.015, respectively, for continuously evaluated dietary and GRS variables). Conclusions: Total fat and especially SFA intake interact with an obesity GRS in determining BMI in a US population. Dietary recommendations to reduce BMI in population with high obesity GRS would be to reduce total fat intake mainly by limiting SFAs.

2164T
Influences of SNPs of LIPC, LIGP, APO E, PLTP and PON1 genes on plasma high-density lipoprotein, cholesterol and paraoxonase-1 activity. D.Z. Scherrer1,1, V.H.S. Zagó 1, R. Secolin 1, E.S. Parra 1, N.B. Panzold 1, F. Alexandre 1, I.V. Calanca 1, E. Nakandakare 1, E.C.R. Quintão 1, E.C. de Faria 1

Background: Our previous studies have shown the determinants of HDL-C levels are under strong control of genes encoding LDL components like lipases, lipid transfer proteins and HDL receptors. The aim of this study was to investigate the influences of several single nucleotide polymorphisms (SNPs) and components of HDL metabolism on plasma levels of HDL-C, cholesterol and paraoxonase-1 (PON1) activity. Material and Methods: Healthy normolipidemic volunteers (Females=159, Males=153; 19 to 75 years of age) were studied. The study was approved by the local Ethics Committee on Research, following the declaration of Helsinki. Genomic DNA was extracted from peripheral blood cells and the SNPs of LIPC (rs1800585, rs2070895), LIGP (rs3813082) APO E (rs4293358, rs7412), PLTP (rs6085904) and PON1 (rs662) detected in the OpenArray® Real Time PCR Platform (Applied Biosystems). Serum Cholesterol and HDL-C were determined by enzymatic methods in an automated system (Hitachi, Roche). The size (nm) and volume (nm³) of HDL were measured in the Nanotrac Particle Size Analyzer (Microtrak, USA). Plasma activities of CETP, PLTP, hepatic lipase and lipoprotein lipase were determined by radioisotopic assays, while PON1 activity by a chromogenic method and endothelial lipase, PLTP mass and plasma insulin were determined by ELISA assays. Lecithin cholesterol acyltransferase (LCAT)- mediated plasma cholesterol was calculated after incubation of plasma samples with LCAT, and endogenous LCAT activity was predicted by plasma cholesterol (R2=0.065; p<0.001), endogenous LCAT activity (R2=0.13; p<0.001), LIPC rs1800585 (TT genotype: R2=0.0025; ps=0.01) and LIPC rs2070895 (GG genotype: R=0.017; ps=0.02). Plasma cholesterol was predicted only by plasma cholesterol (R2=0.12; p<0.0001). Plasma HDL-C was predicted by PON1 rs662 in the TT (R=0.38; p<0.0001) and CT (R=0.19; p<0.001) genotypes. Conclusions: Influences of LIPC rs1800585 and rs2070895 on HDL-C levels were observed in healthy normolipidemic individuals. In addition, PON1 activity was strongly predicted by PON1 rs662 TT and CT genotypes which induced increases in its antioxidant activity. Support: FAPESP (2006/60585-9), CNPq (1599980/2012-7).

2163W
GxE GWAS and path analysis identify a cardiovascular and metabolic risk gene EBF1. A. Singh 1,2,4, M.B. Babyak 1,4, D.K. Nolan 1, B.H. Brunner 1, 2, 3, R. Jiang 3,5, E.C. Siegler 1,4, S.H. Shah 3, 4, R.B. Williams 1,4, E.R. Hauser 2, 3, 5

Background and objectives: Combining multiple genetic variants related to obesity into a global genetic risk score (GRS) might improve identification of individuals at risk of developing obesity. Moreover, characterizing gene-diet interactions is a research challenge aimed at investigating dietary recommendations for individuals with higher predisposition to obesity. The aim was to analyze the association between obesity GRS and body mass index (BMI) in a US population, with focus on gene-diet interactions with total and saturated fat intake. Methods: A cross-sectional study including 783 participants from the Genetics of Lipid Lowering Drugs and Diet Network study. A weighted GRS was calculated on the basis of 62 obesity-associated variants. Anthropometrical and biochemical measurements were taken using standard procedures. Dietary intake was estimated with a validated questionnaire. Results: The obesity GRS was evaluated by tertiles for BMI, participants with a higher GRS had a higher BMI (P trend < 0.001). Participants in the highest GRS tertile and with BMI < 30kg/m2 showed lower total fat (2.5%) and SFA (1.0%) intake compared to participants in the same tertile with BMI ≥ 30kg/m2 (P = 0.010 and 0.027, respectively). An interaction was observed between total fat intake and the obesity GRS for BMI (P for interaction = 0.014 for continuously evaluated dietary and GRS variables). We also identified interactions between SFA and MUFA intake and GRS evaluated continuously for BMI and GRS (P = 0.022 and 0.012, respectively). Finally, we obtained similar results when the SFAs were evaluated categorically according to low and high values (P for interaction = 0.014 for categorical SFA intake). Interactions were also observed between total fat, SFA and MUFA intake and the obesity GRS for waist circumference (P for interaction = 0.004, 0.004 and 0.015, respectively, for continuously evaluated dietary and GRS variables). Conclusions: Total fat and especially SFA intake interact with an obesity GRS in determining BMI in a US population. Dietary recommendations to reduce BMI in population with high obesity GRS would be to reduce total fat intake mainly by limiting SFAs.

2164T
Influences of SNPs of LIPC, LIGP, APO E, PLTP and PON1 genes on plasma high-density lipoprotein, cholesterol and paraoxonase-1 activity. D.Z. Scherrer1,1, V.H.S. Zagó 1, R. Secolin 1, E.S. Parra 1, N.B. Panzold 1, F. Alexandre 1, I.V. Calanca 1, E. Nakandakare 1, E.C.R. Quintão 1, E.C. de Faria 1

Introduction: The determinants of HDL-C levels are under strong control of genes encoding HDL components like lipases, lipid transfer proteins and HDL receptors. The aim of this study was to investigate the influences of several single nucleotide polymorphisms (SNPs) and components of HDL metabolism on plasma levels of HDL-C, cholesterol and paraoxonase-1 (PON1) activity. Material and Methods: Healthy normolipidemic volunteers (Females=159, Males=153; 19 to 75 years of age) were studied. The study was approved by the local Ethics Committee on Research, following the declaration of Helsinki. Genomic DNA was extracted from peripheral blood cells and the SNPs of LIPC (rs1800585, rs2070895), LIGP (rs3813082) APO E (rs4293358, rs7412), PLTP (rs6085904) and PON1 (rs662) detected in the OpenArray® Real Time PCR Platform (Applied Biosystems). Serum Cholesterol and HDL-C were determined by enzymatic methods in an automated system (Hitachi, Roche). The size (nm) and volume (nm³) of HDL were measured in the Nanotrac Particle Size Analyzer (Microtrak, USA). Plasma activities of CETP, PLTP, hepatic lipase and lipoprotein lipase were determined by radioisotopic assays, while PON1 activity by a chromogenic method and endothelial lipase, PLTP mass and plasma insulin were determined by ELISA assays. Lecithin cholesterol acyltransferase (LCAT)- mediated plasma cholesterol was calculated after incubation of plasma samples with LCAT, and endogenous LCAT activity was predicted by plasma cholesterol (R2=0.065; p<0.001), endogenous LCAT activity (R2=0.13; p<0.001), LIPC rs1800585 (TT genotype: R2=0.0025; ps=0.01) and LIPC rs2070895 (GG genotype: R=0.017; ps=0.02). Plasma cholesterol was predicted only by plasma cholesterol (R2=0.12; p<0.0001). Plasma HDL-C was predicted by PON1 rs662 in the TT (R=0.38; p<0.0001) and CT (R=0.19; p<0.001) genotypes. Conclusions: Influences of LIPC rs1800585 and rs2070895 on HDL-C levels were observed in healthy normolipidemic individuals. In addition, PON1 activity was strongly predicted by PON1 rs662 TT and CT genotypes which induced increases in its antioxidant activity. Support: FAPESP (2006/60585-9), CNPq (1599980/2012-7).

To date, several genetic loci and gene networks have been detected for obesity using genome-wide association studies (GWAS) methodology. However, no GWAS study on obesity has been reported for the Saudi population yet. In this study we performed a GWAS for the prevalence of obesity in 5,418 Saudis of Arab ancestry. We discovered 94 genome-wide significant SNPs that reached the threshold of genome-wide significance (p<1x10−5) and may contribute to susceptibility to acquiring obesity. These loci mapped onto or near a number of genes, including the PRE-B-Cell leukemia homeobox 1 (PBX1) and proprotein convertase subtilisin/kexin type 9 (PCSK9) on chromosome (chr) 1, NCK-associated protein 5 on chr 2, caterin (caterin-associated protein), alpha 1 (CNTN1A) and zinc finger RNA binding protein (ZFR) on chr 5, Cub and Sushi domain-containing protein 1 (CSDM1) and zinc fingers and homeoboxes 2 (ZH2X) on chr 8. Other genes mapping on these loci were the V-Rel reticuloendotheliosis viral oncogene homolog A (REL), the olfactory receptor, family 52, subfamily E, member s 1 and 2 (ORS2E1/2), signal-induced proliferation-associated 1 (SIPAI) on chr 11, bicaudal D homolog 1 (bicaudal D) (BICD1) on chr 12, AKAP8 on chr 14, ribosomal protein L15 pseudogene 21 (RPL15P21) on chr 17, chromosome 20 open reading frame 26 (C20orf26) and cadherin 4, type 1, R-cadherin (retinal) (CDHA4) on chr 20. This is the first finding of incident obesity loci identified by GWAS in Saudi Arabs. One other previously reported genetic association, the fat mass and obesity associated (FTO) gene, was replicated, providing support for our study design. These results show a significant correlation of several genomic loci with obesity and provide new insights into pathways contributing to susceptibility for obesity.
Sex-specific effects of CAD SNPs in sudden cardiac death. F.N. Ashar, C. Albert, S.S. Chugh, A. Cupples, M. Eigelheim, P. Goyette, A. Huertas-Vazquez, H. Huikuri, J. Jintila, X. Jouven, S. Kääb, M. Kortelainen, P. Kwok, T. Lehtimäki, L. Lyttyläinen, M. Müller-Nurasyid, C. Newton-Chart, B. Psaty, P. Puill, D. Siscovick, B. Stricker, N. Sotoodehnia, D.E. Arking, CHARGE-SCD. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 2) Division of Preventive Medicine, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts, USA; 3) The Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA; 4) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; 5) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 6) Université de Montréal, Montreal, Quebec, Canada; 7) Dept. of Internal Medicine, University of Oulu and University Central Hospital, Oulu, Finland; 8) Université Paris Descartes, Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Paris, France; 9) Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany; 10) Department of Forensic Medicine, University of Oulu, Oulu, Finland; 11) Department of Dermatology, Cardiovascular Research Institute, and Institute for Human Genetics, University of California, San Francisco, California, USA; 12) Department of Clinical Chemistry, Fimlab Laboratories, School of Medicine, University of Tampere, Finland; 13) Institute of Genetic Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany; 14) Center for Human Genetic Research, Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA; 15) Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA.

Sudden cardiac death (SCD), a major cause of mortality among Western populations, is a clinically heterogeneous condition broadly defined as a sudden unexpected pulseless condition due to a ventricular arrhythmia, while the immediate underlying cause of SCD is considered to be electrical instability, the majority of SCD occurs in the setting of coronary artery disease (CAD). Given this link between CAD and SCD, we chose to investigate the effects of 45 single nucleotide polymorphisms (SNPs) previously associated at genome-wide significance with CAD on their potential association with SCD in a meta-analysis of 4496 cases and over 25,000 controls from 10 studies of European descent. We report a significant enrichment of genetic effects at these SNPs for SCD in the same direction as reported for CAD (32/45, P=0.005). Since both sex and age influence the CAD burden and risk of SCD, we subsequently conducted analyses stratified by sex, as well as limited to samples less than or equal to 65 years of age. As seen in our overall analysis, we continue to observe the enrichment of effects in males (33/45, P=0.002) and younger samples (age=65) (32/45, P=0.005). However, there is complete loss of the enrichment signal in the female-only subgroup (23/45, P=0.88). To investigate whether the enrichment was driven primarily by the presence of underlying CAD, we stratified the analysis on the basis of whether controls were population-based (7 studies) or had been selected to have CAD (2 studies). We confirmed the loss of the enrichment signal in the analysis using the CAD controls (20/45, P=0.54). These findings provide compelling genetic evidence for previous pathological studies that have reported a difference in etiology of disease between males and females, with a greater atherosclerotic burden seen in males than in females, and highlight the importance of assessing genetic associations with SCD stratified by sex.
Blood pressure (BP) is a heritable determinant of risk for cardiovascular disease. To investigate genetic associations with systolic BP (SBP), diastolic BP (DBP), mean arterial pressure (MAP), and pulse pressure (PP), we genotyped ~50,000 single nucleotide polymorphisms (SNPs) that capture variation in ~2,100 candidate genes for cardiovascular disease (CVD) in up to 68,368 individuals of European ancestry from 36 studies. We identified 11 novel associations in independent loci containing 31 genes including PDE1A, HLA-DQB1, CDK6, PRKAG2, VCL, H19, NUCB2, RELA, HOXC6 complex, FBN1 and NFAT5, and 16 previously known associations with SBP, DBP, MAP, or PP, confirmed at the Bonferroni-corrected array-wide significance threshold (P < 6 × 10⁻7) in a combined meta-analysis with an additional independent set of up to 68,968 individuals of European ancestry. An in-depth bioinformatic analysis of the genes and variants in the 11 loci provided functional evidence to support several genes, including cis eQTL associations between SNPs and the expression of HLA-DQB1 and NUCB2. Analysis of ENCODE data identified histone modifications and motifs related to these and other novel BP signals. Drugability analysis for associated genes in this study using existing public resources, including databases of small molecules, shows that ten genes are predicted to be modified by small molecule therapeutics. In summary, we identified novel loci associated with BP and confirmed multiple previously reported associations. Our findings extend our understanding of genes involved in BP regulation, some of which may provide new targets for therapeutic intervention or drug response stratification.

**Posters: Cardiovascular Genetics**

**2174F**

Local ancestry inference in a genome-wide association study of a genetically diverse population with coronary artery disease. Z. Liu, L. Wang, N. Vasudeva, P.J. Goldschmidt-Clermont, M.A. Pericak-Vance, D. Seo, G.W. Beecham. 1. Dr. John T. Macdonald Foundation, Department of Human Genetics, University of Miami, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 3) Division of Cardiology, Miller School of Medicine, University of Miami, Miami, FL.

Coronary artery disease (CAD) is a leading cause of mortality in the United States with an estimated 500,000 deaths each year. Recent studies have shown the prevalence of CAD is different among different racial/ethnic populations with Native Americans being the highest (11.6%), followed by African Americans (6.5%), Hispanics (6.1%), whites (5.8%) and Asian Pacific (3.9%). Genetic risks underlying CAD are not well understood, especially those that are racially/ethnically specific. To study the genetic risk of CAD, we conducted a genome-wide association study (GWAS) of 2,000 catheterization patients from the Miami Cardiovascular Registry of the University of Miami. This dataset represents the diversity of South Florida populations, including European, African, and Hispanic/Latino ancestry. Genotyping was performed on Affymetrix 6.0 platform, and standard QC were performed to ensure data integrity. To address population substructure, we performed local ancestry inference to identify the ancestral state of each locus. Haplotypes were estimated from publicly available reference datasets using ShapeIT. Given phasing uncertainty, we generated 100 phasing iterations for each individual and use them to infer local ancestry with the LAMP-LD/LAMP-ANC software. The 100 iterations were averaged to generate the percentage of ancestry of European, African or Native American at each chromosomal position. For each individual by averaging the results across all chromosomes and comparing it with the estimates derived from a principal component analyses (PCA), Global ancestry estimated by our approach and PCA are highly correlated (r²=0.99; for eigenvector 2 and Native American/Asian, r²=-0.87), suggesting a reliable ancestry estimation by our approach. By summing across individual haplotypes, our samples composed of approximately 75% European, 13% African and 12% Native American/Asian ancestry. We will utilize this local ancestry information to study the contribution of ancestry specific minor alleles to the patient’s anatomic atherosclerotic burden and blood pressure. The advantage of this study is that it allows us to determine common variants associated with cardiovascular disease traits that are common in all populations and unique to a specific race/ethnicity, therefore improving our understanding of CAD genetic risks.
2175W

Genetic variants affecting the expression of DRAM2 at 1p13.3 are associated with acute myocardial infarction with different effects for STEMI and NSTEMI. MI: P. Salo1, J. Sinisalo2, J. Kettunen3, A. Havulinna3, A. Sarin3, T. Hiekkanlina1, S. Ripatti1, P. J. Karhunen1,6, H. Hukkinen4, M. Lokki3, V. Salomaa3, M. Nieminen2, M. Perola1,3,9, 1) Public Health Genomics Unit, Natt Inst Health & Welfare, Helsinki, Finland; 2) Division of Cardiology, Heart and Lung Center HUCH, Helsinki University Central Hospital, Helsinki, Finland; 3) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland; 4) Chronic Disease Epidemiology and Prevention Unit, Natt Inst Health & Welfare, Helsinki, Finland; 5) School of Medicine, University of Tampere, Tampere, Finland; 6) Finlab Laboratories Ltd, Tampere University Hospital Region, Tampere, Finland; 7) Department of Internal Medicine, Oulu University Hospital and University of Oulu; 8) Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland; 9) Estonian Genome Center of the University of Tartu, Tartu, Estonia.

Myocardial infarction (MI) is usually caused by coronary artery disease (CAD). The narrow-sense heritability of MI has been estimated to be 0.35 and 0.38 for males and females. Known genetic risk variants explain 11% of the additive genetic variance in susceptibility to CAD.

Despite advances in understanding the genetic basis of MI, little attention has been paid to the distinction between ST-segment elevation MI (STEMI) and non-ST-segment elevation MI (NSTEMI). The division is made based on characteristic ECG changes and the two are known to present with different characteristics. In most cases STEMI results from a complete occlusion of a coronary artery whereas NSTEMI is caused by partial or transient blockage. NSTEMI tends to present with greater minimal luminal area in the culprit artery and smaller infarct size. Curiously, recurrent infarctions are more often of the same type, suggesting that some individuals may be particularly susceptible to either STEMI or NSTEMI. The possible differences in their genetic risk factors, however, remain virtually unknown.

We performed a genome-wide association study of MI (1579 cases, 1576 controls) with stratification into NSTEMI and STEMI and replicated the results in two independent study samples. We identified a novel risk locus for acute MI at 1p13.11 associated with NSTEMI (OR=1.56, P=4x10^-10) but not with STEMI (OR=1.14, P=0.08). We further show that the SNPs conferring risk for acute MI are also associated with the expression of DRAM2 contained within the NSTEMI-associated region and provide evidence for two distinct but partially overlapping association signals at the locus.

The exact function of DRAM2 remains unknown, but it has been shown to be required for efficient autophagy induction. Silencing DRAM2 inhibits autophagy under starvation and attenuates p53-mediated cell death. The association of variants with both NSTEMI and DRAM2 expression thus suggests regulation of autophagy may affect risk for NSTEMI, raising the possibility that DRAM2 has an effect on the survival of myocardial cells during the ischemic conditions of acute MI. Furthermore, the results also show that genetic factors in part determine whether coronary artery disease results in NSTEMI rather than STEMI.
2177F Genetic Architecture of Hypertension in a Multi-Ethnic Population. N. Vasudeva1, L. Wang1, A. Beecham1, P. Goldschmidt2, M. Penicak-Vance3, D. Seo1, G. Beecham1 1) Hussman Institute for Human Genomics, University of Miami; 2) Department of Medicine, University of Miami.

Previous genetic studies on hypertension have focused mostly on the European ancestry populations. We sought to dissect the genetic basis for hypertension in a diverse population of Hispanics, European ancestry, and African ancestry using a genome-wide association study. The 2,000 samples are ascertained through the clinic-based Miami Cardiovascular Registry, a dataset consisting of ~55% Hispanic and were genotyped using the Affymetrix SNP array 6.0. Quality control tests were performed to ensure data integrity. Linear regression was used to evaluate single variant association with Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Pulse Pressure (PP=SBP-DBP), and Mean Arterial Pressure (MAP=(SBP+2DBP)/3). To account for the medication-lowered blood pressure, we added 7% and 6% for SBP and DBP, respectively, for each antihypertensive medication taken. Age, age2, gender, dyslipidemia and diabetes are significantly associated (P<0.05) with blood pressure in our samples and were included as covariates. Eigenstrat analysis was used to divide samples into Hispanic, White non-Hispanic, and Black subgroups. Analysis was performed within each subgroup and meta-analyzed across subgroups. Strong association (P<0.00001) was found at multiple SNPs for each trait, with the strongest associations at rs9584521 near the MBNL2 gene for PP (P=7.7x10^-8) and rs7278181 in CLDN14 for SBP (P=9x10^-8). Interestingly, both of these loci were monomorphic in the European ancestry subgroup, and would not have shown association in a combined analysis. Additional analyses are underway, including analysis of multiple variants in aggregate and replication in an additional dataset. Our study provides essential data in a diverse population and may lead to additional insights into the genetics of hypertension.

2178W 12 Novel genome wide associations for human cardiac repolarization. N. Verweij1, W.G. Wieringa1, I. Mateo Leach1, D.J. van Veldhuisen1, W.H. van Gilst1, H.L. Hilleges2, R.S.N. Fehrmann2, P.I.W. de Bakker3, R.A. de Boer1, L. Frank2, P. van der Harst1 1) Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands; 2) Trial Coordination Center, University of Groningen, University Medical Center Groningen, Groningen, Netherlands; 3) Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, Netherlands; 4) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands.

Repolarization abnormalities of the heart have been linked to ventricular arrhythmias and sudden cardiac death. Although previous studies of cardiac repolarization have focused largely on interval duration on the electrocardiogram (ECG), changes in the amplitudes may be important as well. Therefore, we performed a GWA meta-analysis on amplitudes of the ST-segment and T-wave (ST-T wave) in 15,943 Dutch individuals. Amplitudes were measured through the lateral, inferior, anterior, septal and AVF leads of the 12 lead electrocardiogram to define the spatial location. We evaluated 2 time points resulting in 10 traits: 5 for the ST segment (at 80ms after J point) and 5 for the T wave. Across 10 traits we identified 41 genotype-phenotype associations (p < 0.25 x 10^-9 based on 8 independent phenotypes) clustered within 17 independent loci. Twelve of the loci are novel determinants for cardiac repolarization. We observed that SNPs with P<6.25E-09 were 3.3 fold enriched for Dnase I hypersensitivity sites in human fetal heart (Z=3.2; compared to 337 other cell types and tissues) and >10 fold enriched for p300 enhancers in the human heart (P<4E-7). This suggests that functional regulatory variants may contribute to several of the associations observed in this study. Furthermore the associations provide insights on the spatiotemporal contribution of genetic variation involved in cardiac repolarization: SNPs at KCND3, TNKS and KCNB1 showed specific associations with the ST-wave. Other loci (XPO1, KCNH2 and TBX3) were specific for the septal leads, possibly related to the function and location of structures such as the AV bundle and its branches. These analyses provide novel insights into the biology of cardiac repolarization and may help to elucidate diseases of cardiac repolarization. Supported by the Netherlands Heart Foundation (grant NHS2010B280).

2179F A three-stage genome-wide association study combining multilocus test and gene expression analysis for young-onset hypertension on Taiwan Han Chinese. K. Chiang1,2,2, H. Yang3, J. Chen3, W. Pan1,2,3 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) Graduate Institute of Life Science, National Defense Medical Center, Taipei, Taiwan; 3) Division of Preventive Medicine and Health Services Research, Institute of Population Health Sciences, National Health Research Institutes, Taiwan; 4) Institute of Biomedical Science, Academia Sinica, Taipei, Taiwan; 5) National Yang-Ming University School of Medical and Taipei Veterans General Hospital, Taipei, Taiwan.

Hypertension is a common and complex disorder. To identify genetic variants will eventually contribute to screening and management efficacy. Although many large-scale genome-wide association studies have been performed, a few studies have successfully identified replicable, large-impact hypertension loci, not to mention the scanty Chinese studies. Young-onset hypertension (YOH) is considered as a more promising target disorder to investigate than the late-onset one due to its stronger genetic component. We performed a three-stage genome-wide association study combining multilocus test and gene expression analysis to map YOH genetic variants. In the first stage, using single locus and multi-locus association tests, we analyzed Illumina HumanHap550 data from 399 YOH cases and 399 age and gender matched controls. In the second stage, differential gene expression analysis was carried out for genes flanked by 14 SNP septets identified in the first stage, using lymphoblast cell lines of the same sample or a joint analysis that combined the data from 1st and 3rd stages. All four are potential novel hypertension susceptibility genes. Among them, the same septet flanking ACTN4 was replicated by Hong Kong Hypertension Study (HKHS) and by WTCCHS hypertension study (WTCCHS), which has been associated with kidney related diseases, such as focal segmental glomerulosclerosis. LARS was replicated in the HKHS, but not in the WTCCHS. GSN which has been implicated in the inflammatory processes in animals may be specific to Taiwanese populations, since it was not validated by HKHS and WTCCHS.
Background: Growth differentiation factor 15 (GDF-15), a cytokine produced in cardiovascular cells under conditions of oxidative stress and inflammation, is an independent prognostic biomarker for cardiac events. Genetic markers located close or in the GDF15 gene have been associated with circulating GDF-15 concentrations. In this genome-wide association study (GWAS) we assess the genetic effects on GDF-15 in a large clinical study on patients with acute coronary syndrome (ACS). Methods: We performed a GWAS of 10 013 ACS patients enrolled in the PLATElet inhibition and aggregation study (PLATO). In total 9448 patients passed genotype quality control and had measurements of GDF-15 levels at baseline. Of these, 3133 patients had additional GDF-15 measurement after one month. Clinical and genetic correlates of GDF-15 were assessed in multivariable analyses. Results: We identified three independent SNPs with genome-wide significance associated with baseline levels of GDF-15; rs17725099 (P-value = 1.47E-107), rs74180880 (P-value = 6.93E-46) and rs1055150 (P-value = 5.89E-10). All three SNPs were in the same region of chromosome 19, close to the GDF15 gene, which has been reported previously to affect levels of GDF-15. The effect on levels of GDF-15 at baseline and at one month was similar for the three SNPs (adjusted beta estimates on log (GDF-15) of 0.08 and 0.16, 0.10 and 0.08, respectively). The identified SNPs are located in predicted regulatory regions, such as an active promoter and strong enhancer, or overlaps with regions predicted to be occupied by transcription factors. Clinical correlates (the top five being age, diabetes, smoking, chronic renal disease and ACS type) accounted for 26.3% of the variation (R2) in log (GDF-15) and the addition of the three top SNPs increased the explained variation by 5.3% to 31.6%. Conclusion: We identified three independent SNPs affecting GDF-15 concentrations. The effect was similar at baseline and at one month follow-up, indicating that the genetic effects on GDF-15 are not restricted to the acute phase of ACS. The large effect on GDF-15 levels by genetic variation might influence the GDF-15 based risk-prediction in patients with ACS. Combining these three SNPs in Mendelian-randomization studies could be used to assess the causal effect of GDF-15 on clinical endpoints in longitudinal cohort studies.
2182T Known SNPs in ADAMTS7, the 9p21 region, ZFAND6 and UBE2E2 interact with type 2 diabetes status to modify the risk of coronary artery disease in large populations. N.R. van Zuydam1,2, B. Voight1, C. Laclaire2, R. Strawbridge3, S. Willems4, E. van Iperen5, J. Hartila6, E. Viachopoulos7, E. Mihaylova8, L. Kwee9, C. Nelson10, M. Kleber11,12, L. Qu1, A. Goel1, J. Kumar5, S. Kanani12, N.W. Rayner12,14, SUMMIT and CARDIOMAPplusC4D. 1) University of Dundee, Dundee, United Kingdom; 2) University of Oxford, Oxford, United Kingdom; 3) University of Pennsylvania, Philadelphia, United States of America; 4) University of Lund, Lund, Sweden; 5) Karolinska Institut, Stockholm, Sweden; 6) Erasmus University Medical Center, Rotterdam, The Netherlands; 7) Academic Medical Center, Amsterdam, The Netherlands; 8) Cleveland Clinic, United States of America; 9) University of Helsinki, Helsinki, Finland; 10) University of Tartu, Tartu, Estonia; 11) Duke University, Durham, United States of America; 12) University of Leicester, Leicester, United Kingdom; 13) LURIC Study, Freiburg im Breisgau, Germany; 14) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 15) University of Uppsala, Uppsala, Sweden.

Patients with type 2 diabetes (T2D) are more likely to suffer from coronary artery disease (CAD) than non-diabetic individuals. In this study we aimed to identify loci that modify the risk of CAD in patients with T2D. We combined summary statistics for 2,295,146 (imputed and directly typed) SNPs from 21,623 patients with T2D (8,456 CAD cases and 13,167 CAD free controls) and 36,104 non-diabetic individuals (14,774 CAD cases and 21,330 CAD free controls) in a fixed effects meta-analysis stratified by T2D: we tested known CAD and T2D loci for an interaction with T2D status and performed a pathway analysis on the meta-analysis summary results. The meta-analysis of allelic effects from CAD case-control associations in patients with T2D identified a signal in ADAMTS7 represented by two independent SNPs previously reported for CAD: rs11072811 (OR=1.2, frequency=0.53, p=8E-12) and rs11634042 (OR=1.1, frequency=0.57, p=4E-08). Rs11072811 had a smaller effect on CAD risk in non-diabetic individuals (OR=1.1, p=1E-02) when compared to its effect in patients with T2D, and this interaction with T2D status was significant (p=9E-03). The meta-analysis of allelic effects from CAD case-control associations in non-diabetic individuals identified the previously published signal in the 9p21 region represented by rs1333042, (OR=1.1, p=5E-02), with a stronger interaction with T2D status in patients with T2D (OR=1.1, p=4E-05). Evaluation of known T2D loci revealed that the T2D risk alleles of rs7612463 (T2D: OR=0.9, P=8E-03; ND: OR=1.1, P=1E-02) in UBE2E2 and rs11634397 (T2D: OR=0.9, P=3E-02; ND: OR=1.1, P=3E-03) in ZFAND6 were associated with decreased risk of CAD in patients with T2D but with increased risk of CAD in non-diabetic individuals (Pinteraction < 5E-03), when compared to the T2D reference allele. Pathway analysis of the combined T2D and non-diabetic CAD associations identified a REACTOME pathway involved in platelet plug formation associated with CAD overall (p=9E-06), while analysis of CAD associations in patients with T2D found associations (p=1E-03) with pathways involved with cell apoptosis (BIOMICARTA), fibroblast growth factor signalling (REACTOME), nitric oxide biosynthesis and phospholipase A2 activity (GOTERM). This study suggests that known CAD SNPs in ADAMTS7 and 9p21, and known T2D SNPs in UBE2E2 and ZFAND6 may differentially modify CAD risk based on T2D status.

2183F Association of Metabochip variants to systolic blood pressure in African Americans from a biorepository linked to de-identified electronic medical records. L. Wiley1,2, R. Goodloe1, E. Farber-Enger1, J. Boston1, D. Crawford1,2, W. Bush1,2,3 Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Molecular Medicine and Biophysics, Vanderbilt University School of Medicine, Nashville, TN.

Blood pressure is a heritable human trait related to a variety of clinical conditions, and hypertensive individuals have increased risk for multiple cardiovascular disorders. The Epidemiologic Architecture for Genes Linked to Environment (EAGLE) BioVU is a subset of the Vanderbilt University biorepository of ~15,000 DNA samples from non-European Americans representing African Americans (n=11,521), Hispanics (n=1,714), and Asians (n=1,122). These samples were genotyped using the Illumina Metabochip to examine influence of these variants on a variety of clinical outcomes. We examined the African American samples to identify genetic variants associated with median outpatient systolic and diastolic blood pressure measurements (excluding patients on blood pressure lowering medications). There were 11,165 patients (63% female) with all available outcomes and covariates. The median BMI of this population was 25, average systolic and diastolic blood pressure were 74.8 (sd=6.5) and 43.9 (sd=4.1) respectively. Median blood pressure measurements were regressed on 192,270 markers from the Illumina Metabochip adjusting for gender, age, body mass index (BMI) and the three most significant principal components to adjust for population stratification. Two variants were associated with systolic blood pressure: rs4945935 and rs10857636. The intragenic variant, rs4945935 (MAF=0.02), was associated with systolic blood pressure (p=3.4e-6). Other variants in this gene have been associated with cardiovascular traits including lipid levels and quantitative subclinical atherosclerosis traits. Additionally this variant is in high linkage disequilibrium (r²=1 in European populations) with a variant that is an eQTL for the gene LRC178, which has been associated with abdominal aortic calcification. Beyond the blood pressure changes and comorbid with multiple clinical conditions, we further adjusted these models for a global composite of all clinical diagnoses. Both associations are robust to this adjustment indicating that these SNPs directly impact systolic blood pressure and are not mediated through association with another condition.

2184W Genome-wide association study of leukocyte telomere length identifies casein kinase 2 (CSNK2A2) to be associated with shorter telomere length and increased cardiovascular disease risk in diabetes. P. Natt1,2, S. Ralhan3, S. Raza1,2, G. Wander3, D. Stowell4, D. Payton4, D. Sanghera5. 1) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Massachusetts General Hospital, Harvard Medical School, Boston, MA; 3) Hero DMC Heart Institute, Ludhiana, Punjab, India; 4) Department of Surgery, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Telomere shortening has been associated with multiple diseases including hypertension, myocardial infarction, osteoporosis, type 2 diabetes (T2D), and Alzheimer’s disease. In this investigation, we have performed a genome-wide scan (GWAS) to identify the common genetic variation that may influence the relative leukocyte telomere length (LTL) and cardiometabolic risk in a diabetic cohort of Punjabi Sikhs from India and the US. Our results revealed a significant association of LTL with coronary heart disease (CHD) the mean LTL showed a gradual decline from healthy subjects to individuals with T2D and CHD showing respective mean LTL being 2.10 in healthy, 1.95 in T2D, 1.69 in CHD, and 1.59 in T2D+CHD. GWAS analysis of discovery cohort (n=1,616/842 T2D cases) identified 338 top independent signals (p<10−4) to be significantly associated with LTL. Most promising 48 SNPs were further replicated through genotyping in an additional Punjabi Sikh sample (n=2,397/1,108 cases). On combined meta-analysis in Sikh populations (n=4,013/1,946 cases), we identified a novel locus in association with LTL at 16q21 represented by an intronic SNP rs74019837 in the CSNK2A2 gene (p = 4.4x10−9). We also found significant association signals near SPATA4, C5orf42, and FER genes for affecting LTL with p values ranging from 2.0x10−7 to 3.3x10−4. Our findings report an independent association of shorter telomere length with T2D and cardiac metabolic risk. Interestingly, telomeric repeat binding factor 1 (TRF1) serves as a substrate for CSNK2A2, which phosphorylates and initiates its binding to telomere. CSNK2A2 also interacts with multiple genes and miRNAs in pathway controlling telomere length and cardiovascular disease. Future functional studies may provide clinically important insights on the interplay of genetic variation in CSNK2A2 and environmental and lifestyle factors for affecting LTL and cardiometabolic risk in diabetes.
A Rho-GTPase pathway related gene is associated with chronic kidney disease via an interaction with coronary artery disease. E. Hauser1, C. Ward-Caviness1, M. Wimm1-2, C. Blach1, C. Haynes1, E. Dowdy2, S. Gregory2, S. Shah3, W. Kraus1-2, 1) Center for Human Genetics, Duke Univ Med Ctr, Durham, NC; 2) Department of Medicine, Duke Univ Med Ctr, Durham, NC.

Chronic kidney disease (CKD) and coronary artery disease (CAD) are intimately linked and often co-occur, implying they may share some aspects of their genetic basis. The genes related to the Rho-GTPase family have been implicated in the pathogenesis of CAD in multiple studies. As regulators of the actin cytoskeleton, Rho-GTPases play an important role in many cellular processes and potentially a function in the pathogenesis of CAD and CKD. Here we present a Rho-GTPase gene that interacts with CAD and is associated with CKD. To discover this association we conducted a genome-wide interaction study in a cardiac catheterization cohort. CATHGEN. We used an ordinal measure of atherosclerosis assessed during the catheterization, CADindex, as a measure of the extent of severity of CAD to test the hypothesis that specific genetic variants (SNPs) that interact with measures of CAD are associated with CKD prevalence CATHGEN. We used race-stratified cohorts of 2202 whites (EA) and 683 blacks (AA) to identify relevant interactions, and a subsequent meta-analysis filtered for consistent associations. We estimated 905,956 SNP-CADindex interactions via a logistic regression model, with CKD as the dependent variable outcome. A Score test was used to calculate P-values, and we adjusted for age, race specific principal components, sex, BMI, hypertension, hyperlipidemia, diabetes, and smoking. We identified the top 10 significant interactions in each cohort, then filtered for those that validated, i.e. P < 0.05 and eSNP-CADindex interaction in each cohort was significant (EA P < 0.05, AA P < 0.05). An EA SNP to validate was rs11014215 located in the 5′ regulatory region of the Rho-GTPase gene ARHGAP21 (EA P = 5.6×10−6, AA P = 0.045, meta-analysis P = 1.3×10−5). ARHGAP21 is a cytoskeletal protein involved in cellular differentiation and is an activator of Cdc42, a cell cycle protein. Examining other interactions in ARHGAP21 revealed one additional association, rs12219181 (ARHGAP21 intron, EA P = 0.032). These analyses use a novel interaction analysis to identify rs11014215, located 5′ of ARHGAP21, as associated with CAD in a SNP-CADindex interaction model. This SNP was strongly associated with CKD in the EA cohort and validated in the AA cohort. Given the previous associations of Rho-GTPase family genes with CAD and the co-occurrence of CAD and CKD we believe this interaction brings novel information to understanding these two important chronic diseases.

91 families. Our GWAS revealed 3 single-nucleotide polymorphisms (SNP)s that were statistically significantly associated with BP near BMPER in Mongolian families. Combining linkage and association enabled the discovery of novel variants associated with BP near BMPER, CA1, and \( \text{CA1} \) were statistically significant after multiple testing corrections (\( P < 5 \times 10^{-8} \)). For fine mapping association test under linkage peaks, we firstly observed two suggestive linkage regions on chromosome 6q27 (LOD for the Rho-GTPase gene ARHGAP21, 21.3 kb away). We also confirmed that the 27 loci identified in previous GWAS were nominally significant (\( P < 5 \times 10^{-8} \)).

High blood pressure (BP) causes global health problem as a major risk factor for cardiovascular disease and death. BP is heritable, with heritability ranging from 18% to 37%. Although recent genome-wide association studies (GWAS) have substantially contributed to understanding the genetic architecture for BP, discovering new variants for explaining missing heritability still remains a major challenge. Moreover, GWAS might miss the true-positive associations within linkage regions because of too stringent significance level for multiple testing. To uncover the novel candidates for BP, we performed fine mapping study under suggestive linkage regions as well as family-based GWAS using an isolated Mongolian sample of 751 individuals from 53 families. Our GWAS revealed 3 single-nucleotide polymorphisms (SNPs) that were statistically significantly associated with BP (\( P < 5 \times 10^{-8} \)). One intergenic SNP near BMPER was moderately associated with BP (\( r = 0.24 \), \( P = 3 \times 10^{-5} \)) and the other two loci were significantly associated with BP (\( r = 0.29 \), \( P = 3 \times 10^{-5} \)).

A family-based linkage and association studies reveal new variants near THBS2 and ACE for blood pressure. H. Kim1, S. Lim1, S. Lee1,2, S. Cho1, J. Sung2, K. Kim3, J. Seo4, 1) Medical Research Center, Genomic Medicine Institute (GMI), Seoul National University; 2) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea; 3) Seoul National University School of Public Health, Seoul 151-742, Korea; 4) Psoma Therapeutics Inc., Seoul 153-781, Korea; 5) Macrogen Inc., Seoul 153-781, Korea.

Posters: Cardiovascular Genetics

For fine mapping association test under linkage peaks, we firstly observed two suggestive linkage regions on chromosome 6p27 (LOD for \( SBP = 2.4 \)) and \( 5q35 \) (LOD for \( DBP = 2.0 \)), which were consistent with previously reported linkage peaks in other populations. The subsequent association test in presence of linkage found 7 and 2 SNPs in each linkage region, statistically significantly \( (P < 5 \times 10^{-8}) \). The strongest linkage peak was found for rs9393162, an intergenic SNP near THBS2 (21.3 kb away). We also confirmed that the 27 loci identified in previous GWAS were nominally replicated \( (P < 0.05) \), and of these loci, ADM and ZNF652 for \( SBP = 2.0 \) and \( DBP = 2.5 \). Whole genome linkage and association highlighted the importance of the 5q35 region, which included several candidate genes such as \( ACE \), \( BMPER \), and \( IL5 \). Of these, rs4459609, an intergenic SNP near \( ACE \) gene (5.5 Kb away) showed a significant effect on both \( SBP = 4.75 \times 10^{-5} \) and \( DBP = 8.59 \times 10^{-5} \).
2189F

Background: Red cell alloimmunization in transfused sickle cell disease patients occurs in approximately 30%. Because red blood cell (RBC) transfusion is an important part of therapy, the need for additional antigen matching once alloimmunization occurs is problematic, and leads to therapeutic limitations. Thus, it would be important to identify risk factors for alloimmunization in this patient population. We performed a genome wide association study (GWAS) in an attempt to identify possible genes associated with RBC alloimmunization. Methods: 157 sickle cell disease patients, including 69 who developed alloimmunization and 88 who did not were genotyped for 2,217,402 SNPs using the Affymetrix PanAFR Array. SNPs with a call rate >90%, Hardy Weinberg Equilibrium P=0.0001, or having a minor allele frequency (MAF) <1% were excluded from the analysis. In total, 2,113,177 SNPs were analyzed. Population structure was examined by principal components analysis (PCA) using the GWAS SNPs. Statistical analyses were performed using Golden Helix SVS 7. To correct for multiple testing, the genome-wide significance threshold was set at P<2.4 × 10^-8.

Of note, the higher WAA were more likely to develop RBC alloantibodies. A larger sample size would be necessary to find SNPs which reach genome-wide significance (P=1.31 × 10^-6; 2.82 × 10^-6; and 3.21 × 10^-6, respectively).

2190W
Identifying population specific dyslipidemia variants using cross-population GWAS. A. Koivu,1 D. Weissglas-Volkov,4 R.M. Cantor,1 E. Nikkola,4 K.A. Deere,1 J.S. Sinsheimer,1 B. Pasanen,2 R. Brown,2 V. Salomaa2, J. Kaprio,5,6, A. Loukola,6 A. Jula,4 M. Jauhiainen,4 M. Heliovaara4, O. Raitakari,7,8 T. Lehtimaki,7,9, J.G. Eriksson,4,10,11 M. Perola11, L.B. Riba12,13 T. Tusie-Luna12,13, C.A. Aguilar-Salinas11, P. Paikukanta1,1 Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) 2Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA; 3) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 6) Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere, Finland; 10) Folkhälsan Research Center, University of Helsinki, Helsinki, Finland; 11) Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; 12) Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico; 13) Instituto de Investigaciones Biomédicas de la UNAM, Mexico City, Mexico.

The high prevalence of dyslipidemia in Mexicans is a serious health burden due to the increased risk for cardiovascular disease and medical cost. Identifying Hispanic-specific variants can accelerate our understanding of lipid disease pathogenesis in this rapidly growing U.S. minority. However, studies in admixed populations are hindered by complex population substructure that hampers power. To address this, we could leverage the cross-population GWAS approach, which may otherwise be missed by conventional GWAS.

Identifying populationspecific dyslipidemia variants using cross-population GWAS. The high prevalence of dyslipidemia in Mexicans is a serious health burden due to the increased risk for cardiovascular disease and medical cost. Identifying Hispanic-specific variants can accelerate our understanding of lipid disease pathogenesis in this rapidly growing U.S. minority. However, studies in admixed populations are hindered by complex population substructure that hampers power. To address this, we could leverage the cross-population GWAS approach, which may otherwise be missed by conventional GWAS.

Identifying population specific dyslipidemia variants using cross-population GWAS. A. Koivu,1 D. Weissglas-Volkov,4 R.M. Cantor,1 E. Nikkola,4 K.A. Deere,1 J.S. Sinsheimer,1 B. Pasanen,2 R. Brown,2 V. Salomaa2, J. Kaprio,5,6, A. Loukola,6 A. Jula,4 M. Jauhiainen,4 M. Heliovaara4, O. Raitakari,7,8 T. Lehtimaki,7,9, J.G. Eriksson,4,10,11 M. Perola11, L.B. Riba12,13 T. Tusie-Luna12,13, C.A. Aguilar-Salinas11, P. Paikukanta1,1 Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) 2Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA; 3) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 6) Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere, Finland; 10) Folkhälsan Research Center, University of Helsinki, Helsinki, Finland; 11) Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; 12) Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico; 13) Instituto de Investigaciones Biomédicas de la UNAM, Mexico City, Mexico.

The high prevalence of dyslipidemia in Mexicans is a serious health burden due to the increased risk for cardiovascular disease and medical cost. Identifying Hispanic-specific variants can accelerate our understanding of lipid disease pathogenesis in this rapidly growing U.S. minority. However, studies in admixed populations are hindered by complex population substructure that hampers power. To address this, we could leverage the cross-population GWAS approach, which may otherwise be missed by conventional GWAS.
Arteriosclerosis obliterans (ASO) is the most common cause of peripheral vascular disease affecting the lower limbs. ASO mainly occurs from atherosclerosis, resulting in obstruction of the blood supply to the lower or upper extremities. ASO might result from complex interactions multiple genetic and environmental factors. To reveal genetic backgrounds in the pathogenesis for ASO, we performed a genome wide association study (GWAS) in a Japanese population. Genotyping was performed by the Illumina HumanHap550v3 for controls. We applied stringent quality-control criteria and tested 789 cases and 3,383 controls for 510,687 autosomal SNPs commonly available on both BeadChip. The inflation of test statistics, \( \hat{\lambda} \), was 1.04. Through combination of this GWAS and a following staged analysis with a total of 2,647 Japanese subjects with ASO and 20,560 control subjects, we have identified two susceptible loci for this disorder on chromosome 13 and 4 (P values < 10^-5). To our knowledge, these are the first genetic risk factors identified for ASO.

Common autosomal variants are associated with bicuspid aortic valve in Turner Syndrome. S.K. Prakash1, M. Silberbach2, S. Hooker3, D.C. Guo4, C. Masen1, C.A. Bondy1, D.M. Milewicz4, GenTAC Investigators. 1) University of Texas Health Science Center at Houston, Houston, TX; 2) Oregon Health & Science University, Portland, OR; 3) Baylor College of Medicine, Houston, TX; 4) National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD.

Background: The prevalence of bicuspid aortic valves (BAV) is enriched thirty-fold in women with Turner Syndrome (TS) in comparison with the general population. We hypothesize that common autosomal variants influence the development of BAV in TS women, who may be uniquely sensitized to these variants by the loss of one X chromosome. We sought to identify autosomal BAV susceptibility genes in a cohort of TS women (average age 33 years, 38% BAV, 15% coarctation). Methods: 73 TS women of European ancestry with BAV and 120 TS women with tricuspid aortic valves were genotyped using Illumina Omni-Express arrays (668,262 SNPs). Tests of association were performed using logistic regression without adjustment for covariates. Results: TS women with 45,X karyotypes were not significantly more likely to have a BAV than those with isochromosomes, rings or Xp deletions (P=0.09). After removal of 11 outlier samples for the first two multivariable scaling components, the genomic inflation factor was 1. The strongest association signals were observed on chromosomes 2, 5 and 22 and did not overlap with previously reported loci for BAV. A total of 16 SNPs in the ITGA1 locus on squ112 were positively associated with BAV (OR=4.3) with a minimum \( P \) value of 1.5x10^-5. ITGA1, which encodes the alpha-1 integrin, is a promising BAV candidate gene because it is expressed in valve tissue and is required for vascular cell adhesion and myofibroblast differentiation. Replication of these regions in independent groups of cases is ongoing. Conclusion: This is the first study to demonstrate that autosomal variants are associated with BAV in TS women, and provides evidence for gene-gene interactions in BAV formation.

Using common genetic variants to predict dyslipidemia. C.M. van Duijn1, S. Willems2, A. De Waal3, B.A. Oostra4, O.H. Franco1, A. Isaacs3. 1) Dep Epidemiology, ErasmusMC, Rotterdam, Netherlands; 2) Department Internal Medicine, ErasmusMC, Rotterdam, Netherlands; 3) Department Clinical Genetics, ErasmusMC, Rotterdam, Netherlands.

Hypercholesterolemia (HC) is an important modifiable cardiovascular disease risk factor. We determined the ability of genotypic risk scores to identify individuals at increased risk of hypercholesterolemia. We calculated TC, LDL-C, HDL-C, and TG risk scores in an additional 6,688 samples from the Tromsø study (Tromsø 4) including 2,350 MI cases and 2,318 controls, all from Norway. We evaluated the array coverage in these Norwegian samples using whole genome sequencing combined with exome enrichment in 162 samples. The exome array provided successful genotyping for an estimated 74% of Norwegian loss-of-function or missense variants with frequency > 5%, 73% of coding variants with frequency 1-5% and 51% of rare coding variants with frequency < 1%. Despite sufficient power to detect association with low frequency variants (1-5% frequency) with moderate to high effect sizes (odds ratio 1.38-1.91) in the combined sample size of 10,333 individuals, we did not identify any novel genes or single variants that reached significance. However, we identified several strong candidate genes hovering below significant effect sizes suggesting that low frequency and rare variants with intermediate effect sizes likely will be identified with larger samples.

2193W

Exome-wide Coding Variation and Myocardial Infarction. H. Zhang1, O.L. Holmen12, E. Schmidt3,4, M. Lochen5, C. Platou6, E.B. Mathiessen4, T. Helledam6, M. Christensen6, J.D. Mikkelsen6, I. Njolstad9, K. Hveem12, C.J. Willer1,7. 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway; 3) Laboratory for Genotyping Development, RIKEN, Ctr Integrative Medical Science.; 4) Laboratory for Genotyping Development, RIKEN, Ctr Integrative Medical Science.; 5) Human Genome Center, Institute of Medical Science, the University of Tokyo, Tokyo, Japan; 6) Section of Hematology/Oncology, The University of Chicago; 7) Department of Human Genetics and Disease Diver- sity, Tokyo Medical and Dental University.

The prevalence of bicuspid aortic valves (BAV) is enriched thirty-fold in women with Turner Syndrome, and did not overlap with previously reported loci for BAV. A total of 16 common autosomal SNPs commonly available on both BeadChip. The inflation of test statistics, \( \hat{\lambda} \), was 1.04. Through combination of this GWAS and a following staged analysis with a total of 2,647 Japanese subjects with ASO and 20,560 control subjects, we have identified two susceptible loci for this disorder on chromosome 13 and 4 (P values < 10^-5). To our knowledge, these are the first genetic risk factors identified for ASO.

Using common genetic variants to predict dyslipidemia.
2195F
Asian Institute for Life Sciences, University of Ulsan College of Medicine, Seoul, South Korea.

Kawasaki disease (KD) is an acute self-limited vasculitis of infants and children, manifested by fever and signs of mucocutaneous inflammation. Children with KD show laboratory abnormalities of inflammation, such as elevated white blood cell (WBC) count, C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), and they also may have anemia, thrombocytosis, hypoalbuminemia, and elevated serum transaminases, etc. To identify genetic loci influencing important laboratory markers in KD - WBC count, neutrophil count, platelet count, CRP, ESR, hemoglobin (Hb), aspartate aminotransferase (AST), alanine aminotransferase (ALT), albumin, and total protein, we performed association studies using our previous genome-wide association study (GWAS) data of KD. Linear regression analyses after adjustment for age and sex were carried out for the 10 quantitative traits in 178 KD patients. A total of 165 loci passed our arbitrary stage 1 threshold for replication ($P < 1 \times 10^{-8}$). For CRP, variants in the CRP gene showed the most significant associations ($P = 1.20 \times 10^{-6} - 5.69 \times 10^{-6}$) and for WBC count, a variant in the TNFRSF1B gene reached genome-wide significance level ($P = 2.80 \times 10^{-6}$). For HB, variants in the SOX5 gene were the most significantly associated ($P = 9.40 \times 10^{-7} - 4.41 \times 10^{-6}$) and for albumin, the most compelling association involved a variant in the CDKAL1 gene which was reported as being associated with albuminuria ($P < 8.73 \times 10^{-6}$). Of the significant variants from the stage 1 GWAS analysis, 22 were selected as candidates considering their functional importance. Further replication studies are planned to validate these results.

2196W
Novel association of endothelial function with a variant in PEAR1. A.S. Fisch1, P. Donnelly1, M. Drote1, S. Newcomer1, A. Panhari1, K.A. Ryan1, W. Herzog2, A.R. Shuldiner2, J.P. Lewis2.
1) Program in Personalized and Genomic Medicine, and the Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 2) Department of Medicine, Johns Hopkins Medical Institute, Baltimore, MD.

The anti-platelet agent aspirin is one of the most commonly prescribed medications for prevention and treatment of coronary heart disease. However, there is great inter-individual variation in response to aspirin as measured by on-treatment residual platelet aggregation and cardiovascular outcomes. In the Pharmacogenomics of Anti-Platelet Intervention (PAPI) Study, a genome-wide association study (GWAS) in 670 healthy Old Order Amish (OOA) participants, we identified a genome-wide significant association between on-aspirin collagen-stimulated platelet aggregation and PEAR1 genotype (rs12041331). The rs12041331 variant was also associated with cardiovascular event risk in two independent populations. PEAR1 (platelet endothelial aggregation receptor 1) is a platelet-platelet contact receptor that modulates platelet aggregation. In addition to platelets, PEAR1 is highly expressed on endothelial cells, which led us to hypothesize that the variant’s association with cardiovascular risk may be mediated through effects on endothelial function. To investigate this hypothesis, we genotyped rs12041331 in participants in the Heredity And Phenotype Intervention (HAPI) Heart Study, a clinical intervention study with 868 healthy OOA participants in whom cardiovascular traits were assessed, including platelet aggregation and endothelial function by brachial artery flow-mediated dilation (FMD). In this cohort, we confirmed significant association between on-aspirin collagen-stimulated platelet aggregation and PEAR1 genotype (effect size = 6.56, $p = 0.037$), with the minor A allele associated with decreased platelet aggregation. In addition, we found significant association between PEAR1 genotype and FMD (effect size = 1.14, $p = 0.039$), with the A allele associated with increased FMD, which represents greater endothelial function. These results indicate that PEAR1 likely plays a dual functional role in platelets and endothelium, and that PEAR1 rs12041331 genotype is associated with both. Future studies will be required to more precisely understand the mechanisms by which PEAR1 and its genetic variants influence aspirin response, ultimately leading to the development of more individualized and effective anti-platelet therapy.
2198F

PMA (phorbol 12-myristate 13-acetate) regulates in vitro alternative splicing of ORL1, a gene involved in atherosclerosis and tumorigenesis. E. Marinò1, J.R. Tejedor2, B. Rizzaccas3, M.C. Belloccchi4, F. Ferré1, A. Botta4, D. Caporossi1, J. Valcârce5, G. Novelli6,7, F. Amati1. 1) Dept. of Movement, Humanities and Health Sciences, “Foro Italico” University, Rome, RM, Italy; 2) Dept. of Biomedicine and Prevention, Tor Vergata University, Rome, RM, Italy; 3) Centre de Regulació Genòmica and Universitat Pompeu Fabra, Barcelona, Spain; 4) Dept. of Experimental Medicine and Surgery, Tor Vergata University, Rome, RM, Italy; 5) Dept. of Biology, Tor Vergata University, Rome, RM, Italy; 6) Dept. of Internal Medicine, University of Arkansas for Medical Sciences and Central Arkansas Veterans Healthcare System, Little Rock, AR, USA; 7) St. Peter Fatebenefratelli Hospital, Rome, RM, Italy.

Lectin-like oxidized low-density lipoprotein receptor-1 (LOX-1), encoded by the ORL1 gene, is the major endothelial receptor for ox-LDL and plays a fundamental role in the pathogenesis of atherosclerosis. Furthermore recently, it has been demonstrated that an increased activity of LOX-1 is associated with cancer cell invasion. ORL1 is subjected to a physiological alternative splicing, if its isoform, Loxin, lacks exon 5 and encodes for a putative truncated receptor (LOXIN) with impaired binding activity. LOXIN is considered as a natural inhibitor of LOX-1 mediated signalling and each effort to increase LOXIN expression may have a potential therapeutic SNP. The expression of ORL1 and Loxin is regulated by six intrinsic SNPs in linkage disequilibrium and the H-risk haplotype is associated to acute myocardial infarction (AMI) and atherosclerosis. The macrophages of individuals with H-risk haplotype present a ORL1/Loxin mRNA ratio 33% higher compared to individuals with L-risk haplotype. In order to study in detail the regulation of ORL1/Loxin, we have analyzed the expression of the two isoforms in vitro, leading to an increase of the less functional Loxin isoform.

2199W

Identification of a predictive/prognostic genetic signature in Chagas Cardiomyopathy: A systems biology approach on the site of action. C. Chevillard1,2, J.R. Tejedor1, B. Rizzaccas2, M.C. Belloccchi4, F. Ferré1, A. Botta4, D. Caporossi1, J. Valcârce5, G. Novelli6,7, F. Amati1.

Although recent genome-wide association studies (GWAS) have identified a handful of variants with best significance for coronary artery disease (CAD), it remains a challenge to summarize the underlying biological information from the abundant genotyping data. Here, we propose an integrated network analysis that effectively combines GWAS genotyping dataset, protein-protein interaction (PPI) database, literature and pathway annotation information. This three-step approach was illustrated for a comprehensive network analysis of CAD as the followings. First, a network was constructed from PPI database and CAD seed genes mined from the available literatures. Then, susceptibility network modules were captured from the results of the network analysis identified four susceptibility modules for CAD including a complex module consisting of 15 functional interconnected sub-modules, AGPAT3-AGPAT4-PPAP2B module, ITGA11-ITGB1 module and EMCN-SELL module. MAPK10 and COL4A2 among the top-scored focal adhesion pathway related module were the most significant genes (MAPK10: OR = 32.5, P = 3.5x10-11; COL4A2: OR = 2.7, P = 2.8x10-10). The significance of the two genes were further validated by other two gene-based association tests (MAPK10: P = 0.009 and 0.007; COL4A2: P = 0.001 and 0.023). The susceptibility modules identified in our study might provide novel clues for the clarification of CAD pathogenesis in the future.
2201F
Systematic phenotype prediction in zebrafish identifies novel, disease-relevant cardiovascular gene functions. G. Musso1,2, M. Tasan1,2, C. Mosimann1,4,6,7, J.E. Beaver3, E. Plojiev4, L.A. Carr1,2,6,7, H.N. Chua4, J. Dunham4, K. Zuben4, Q. Morris4, L. Zorn5,6, F.P. Roth3,4,6,8, C.A. MacRae1,2. 1) Department of Medicine, Harvard Medical School, Boston, MA; 2) Cardiovascular Division, Brigham and Women’s Hospital, Boston, MA; 3) Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA; 4) Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, ON; 5) Howard Hughes Medical Institute, Boston, MA; 6) Stem Cell Program, Children’s Hospital Boston, Boston, MA; 7) Division of Hematology/Oncology, Children’s Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School, Boston, MA; 8) Samuel Lunenfeld Research Institute, Mt. Sinai Hospital, Toronto, ON.

Comprehensive functional annotation of vertebrate genomes is a central step in realizing the full potential of genome-scale technologies for fundamental and translational discovery. However, the identification of novel gene functions through forward genetic screening quickly becomes overwhelming not only due to the number of surveyable genes, but also due to the increasing number of observable phenotypes. Unbiased prediction of gene-phenotype relationships offers a strategy to direct finite experimental resources towards likely phenotypes, thus maximizing de novo discovery of gene functions. Here we prioritized genes for phenotypic assay in embryonic zebrafish through genomic data integration and machine learning, predicting the effect of loss of function of each of 15,106 zebrafish genes on 338 distinct anatomical processes. Cross-validation suggested the resulting predictions to be particularly precise for certain phenotypic categories, including cardiac phenotypes. In proof-of-concept studies we validated 16 high-confidence cardiac predictions using targeted morpholino knockdown and manual blinded phenotyping in embryonic zebrafish, confirming a significant enrichment for cardiac phenotypes as compared to morpholino controls. Subsequent detailed analyses of cardiac function confirmed these results, identifying novel physiological defects for 11 tested genes, the majority of which are conserved in humans. Among these we identified tmem88a, a recently described attenuator of Wnt signaling, as a discrete regulator of the patterning of heart field progenitors. The results suggest that large-scale gene prioritization in zebrafish enriches for phenotype association, thus substantially reducing survey effort in identifying novel, disease-relevant vertebrate gene functions.

2202W
Large-scale transcriptome profiling in peripheral blood mononuclear cells (PBMCs) of early-onset myocardial infarction individuals and matched controls. C. Müller1,2, A. Schott1,2, M.O. Scheinhardt1, S. Schulthess1, M.O. Scheinhardt, S. Szymboczak2, F. Ojeda1, C.R. Sinning1, R. B. Schnabel1, S. Wilde1, P.S. Wild4, K.J. Lackner2, T. Munzel2, A. Ziegler2, S. Blankenberg1, T. Zeller1. 1) Clinic for general and interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany; 2) Institute of Medical Biometry and Statistics, University Hospital Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; 3) Clinical Epidemiology, Center for Thrombosis and Haemostasis, University Medical Center Mainz, Mainz, Germany; 4) Department of Medicine II, University Medical Center Mainz, Mainz, Germany; 5) Institute for Clinical Chemistry and Laboratory Medicine, University Medical Center Mainz, Mainz, Germany.

Despite the substantial progress in diagnosis and therapy, myocardial infarction (MI) remains one of the major causes of mortality worldwide. MI in young adults is a rare phenomenon and the molecular mechanisms of the pathogenesis leading to MI at young age are still unclear. The aims of this study were to identify differentially expressed genes between young, non-acute MI individuals and healthy controls and to highlight underlying biological mechanisms by functional over-representation analyses. In 112 non-acute MI individuals (age at MI <= 50) and 112 age and sex matched healthy controls from the Gutenberg Health Study (GHS) gene expression levels in PBMCs were determined using Affymetrix GeneChip Exon ST1 Arrays. Differential gene expression was computed by applying linear mixed models adjusted for body mass index, hypertension, smoking status, diabeties and LDL/HDL ratio and the first four principal components. Functional analyses of differentially significantly expressed genes (FDR<=0.1) were conducted using Ingenuity Pathways. As a result of our analysis, gene therapy and gene expression were identified in PBMCs of GHS controls (n=307). 183 genes were significantly associated with MI. The top four differentially expressed genes were G-protein-coupled receptor 15 (GPR15, up-regulated), ATP-binding cassette transporter A1 (ABCA1, down-regulated), selenoprotein O (SELO, down-regulated) and cardioprotein 1 (CTF1, up-regulated). We observed a strong up-regulation of GPR15 with smoking behaviour, but the association with MI remained significant after adjustment. It had been shown that the GPR15 locus in leukocytes of smokers is hypomethylated, indicating a higher accessibility of the DNA and subsequent increased expression of GPR15, which is reflected in our data. Moreover, we observed a positive correlation between GPR15 expression and plaque formation. Pathway analysis revealed over-representations of differentially expressed genes in key pathways of the cardiovascular system. The top gene GPR15 shows strong associations to smoking and plaques. Ongoing functional analysis will help to understand its role in cardiovascular disease, which could be promoted by smoking.

2203T
Investigating the role of HoxA3 during cardiac development. V. Sanghvi1, D. Rux1, S. Chan2, L. Borges3, N. Koyano4, R. Perlinger5, D. Garry3, M. Kyba3, M. Iacovino1. 1) Pediatric, Medical Genetics, LABiomed Research Institute at Harbor UCLA, Torrance, CA; 2) Department of Pediatrics, Lillehei Heart Institute, University of Minnesota, MN; 3) Department of Medicine, Lillehei Heart Institute, University of Minnesota, MN.

During embryoid body (EB) differentiation, EB differentiation mesoderm patterns toward its major derivatives: lateral plate mesoderm to generate blood and vasculature and paraxial mesoderm to generate muscle and cardiac derivatives. Day 4 EB differentiation thus resembles the gastrulating embryos at EPC 7 (Days Post Conception). Previous studies performed by our group showed a role of HoxA3 on the generation of hematopoietic progenitor cells from haemogenic endothelium, the progenitor cell of Hematopoietic Stem Cell (HSC) in vivo. As a result of HoxA3 upregulation the genome transcriptional profile showed a repression of hematopoietic commitment and promotion of endothelial fate. Unexpectedly we found a strong repression of cardiac specific regulators both in endothelial-committed progenitors and total EBs. HoxA3 gene expression results to be crucial during developmental stage, indeed genetic deletion in mice results in perinatal lethality due to cardiovascular outflow tract defects as well as pulmonary defects and, as an overall phenotype, in recapitulate DiGeorge syndrome. Fate mapping has shown contribution of HoxA3 expressing cells to the outflow tract as well as atria and ventricle. We demonstrate with functional assays and at molecular level, that upregulation of HoxA3 in mesoderm patterning strongly represses both cardiac commitment and differentiation. We are dissecting the molecular mechanism underlying this repression and we show preliminary evidence that HoxA3-dependent cardiac repression is mediated through regulation of the Notch ligand, Jagged1.
Joint Association of 31 Mitochondrial Variants with Type 2 Diabetes: The Strong Heart Family Study, Y. Zhu1, ET. Lee2, SA. Cole3, K. Haack4, LG. Best5, BV. Howard6, J. Zhao1. 1) Dept Epidemiology, Tulane Univ Pub Health & Tropical Med, New Orleans, LA; 2) Center for American Indian Health Research, University of Oklahoma Health Sciences Center; 3) Texas Biomedical Research Institute, San Antonio, TX; 4) Missouri Breaks Industries Research Inc, Timber Lake, SD.

Background: Multiple mitochondrial variants have been associated with T2D, but a single SNP usually confers a small risk to disease. A gene-family approach taking into account the joint contribution of multiple variants may have greater power in detecting genetic associations. Objective: To evaluate the cumulative impact of 31 tagging SNPs in 13 mitochondria-related genes in American Indians participating in the Strong Heart Study (SHS). Methods: Thirty-one tagging SNPs in thirteen mitochondrial-related genes were genotyped in 1,221 American Indians residing in South/North Dakota. All subjects were recruited and examined in 1989-1991 by the SHS. We first conducted single SNP analysis by logistic regression, adjusting for age, sex, BMI, socioeconomic status, lifestyle factors, (physical activity, alcohol intake and smoking status), estimated glomerular filtration rate (eGFR) and history of cardiovascular disease or hypertension. Gene-based association was assessed by combining p-values of all SNPs within a gene based on single SNP analysis using a weighted truncated product method. The gene-family association was performed by combining p-values of all 13 genes using the same method. Sensitivity analysis was conducted to evaluate whether the observed associations were primarily driven by the most significant SNPs. Multiple testing was corrected using false discover rate (FDR). Results: Multiple SNPs showed nominal or marginal associaton with T2D or insulin resistance, but the association diminished after correction for multiple testing. Four genes, including CYTB (p=0.04), COX3 (p=0.04), ND1 (p=0.037) and TRNC (p=0.018), were weakly associated with T2D, but the associations disappeared after correction for multiple testing. However, a gene-family analysis comprising all 31 SNPs from the 13 genes showed significant association with T2D, suggesting that mitochondrial alleles defined by haplotype of these SNPs contribute to T2D susceptibility. This association was not driven by the most significant SNPs. Conclusion: Although a single SNP confers a small risk to T2D, multiple variants in the mitochondrial pathway may jointly contribute to T2D, suggesting the importance of modeling the joint impact of multiple variants on complex disease susceptibility. Our findings may provide useful information for risk classification or personalized approach to diabetes management in American Indians.
2206T
Genes related to CRKL and mouse heart development may act as genetic modifiers to congenital heart disease in human 22q11.2 deletion syndrome. J. Chong1, S.E. Racedo1, T. Guo1; D.M. McDonal1, McGinn1, 2, E. Zackai2, 3, R.J. Shprintzen4, B.S. Emanuel1, 2, B. Funke5, 6, 7. B.E. Morrow1. 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Division of Human Genetics, The Children’s Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 4) The Virtual Center for Velo-Cardio-Facial Syndrome, Inc., Manlius, NY; 5) Laboratory for Molecular Medicine, Partners Center for Personalized Genetic Medicine, Cambridge, MA; 6) Department of Pathology, Massachusetts General Hospital, Boston, MA.

Congenital heart disease (CHD) occurs at a rate of 7-10:1000 live births and 80% of CHD cases are of unknown etiology. Approximately 70% of patients with 22q11.2 deletion syndrome (22q11DS), also known as DiGeorge/VCFs, OMIM: 192430; 188400) have a CHD, mostly of the conotruncal type, with highly variable expressivity suggesting the presence of genetic modifiers. The majority (~85%) of patients with 22q11DS have a 3 Mb deletion encompassing ~60 genes including the T-box transcription factor, TBX1 (OMIM: 602054) and the adaptor protein, CRKL (OMIM: 602007), both implicated in the etiology of CHD in the disorder. There have been extensive studies regarding CHD on TBX1, but few have been performed on CRKL, which links various signaling pathways to downstream effectors. To elucidate the independent role of CRKL in human CHD, we analyzed echocardiograms from 15 subjects with 22q11DS encompassing CRKL but not TBX1. We found 7/15 have cardiac outflow tract malalignment or septal defects, which are strikingly similar to those seen in 22q11DS. We will test for genetic interactions between CRKL and other genes harboring an allelic series of Crkl dosage, suggesting that CRKL is important in its own right, and that it may serve as a genetic modifier of CHD in subjects with the typical 3 Mb deletion.

To follow up on our observations, we performed a candidate modifier gene test of 22q11DS and identify additional modifiers related to CRKL, we performed whole exome sequencing (WES) on 176 subjects with the typical 3 Mb deletion; 74 subjects have tetralogy of Fallot and 102 have a normal heart/aortic arch. We performed a candidate gene association test, using SKAT (PMID: 23684098), on 141 genes either in CRKL pathways or with a known role in the development of structures related to phenotypes in Crkl mutant mice. Of the 141 genes, 103 have nonsynonymous single nucleotide variants (SNVs). Six genes (NFATCD4, HBEGF, CRKL, HSPG2, PXN, and ELN/MST2) have a nominal p-value < 0.05 and empirical p-value < 0.1 after 10,000 permutations. Our top gene is NFATC4 (p-value = 0.0073, FDR q-value = 0.2509, OMIM: 602699). Importantly, mice with mutations in Nfatc4 have similar heart malformations as our Crkl mutant mice indicating that nonsynonymous variants in NFATC4 may genetically interact with CRKL to modify the CHD phenotype seen in 22q11DS. We will test for genetic interactions between Crkl and Nfatc4 using mouse models and perform replication studies of our findings in an independent cohort of 22q11DS subjects.

2207F
Genetic Influence of Scavenger Receptor Class B Type 1 (SCARB1) on Plasma Lipid Traits in non-Hispanic White Americans. V. Niemzun1, 2, X. Wang1, 2, J.E. Hoke1, 2, J.部位1, 2, 3, M. M. Kamboun1, 2, 3, 4. 1) Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO.

Coronary artery disease (CHD) is a major public health burden worldwide. A main risk factor for CHD includes low levels of high-density lipoprotein cholesterol (HDL-C). SCARB1 is one of the candidate genes involved in HDL-C metabolism. To identify genetic influences of both common and rare variants of SCARB1 on plasma lipid traits, we sequenced all 13 exons, exon-intron boundaries plus 1 kb of 5’ and 3’ flanking regions of SCARB1 in 95 non-Hispanic White (NHW) US individuals with the upper (n = 47) and lower (n = 48) 10th percentile of HDL-C/triglyceride (TG) distribution. Sequencing and cell-based functional assays identified 44 variants (MAF ≥5%, n = 12; MAF <5%, n = 32). No statistically significant difference was observed in the distribution of rare variants between the two extreme HDL-TG groups (P = 0.615). We genotyped 40 relevant common variants and common tag SNPs identified by our sequencing (MAF ≥5%, n = 8; MAF <5%, n = 32), plus 32 additional tag SNPs from HapMap covering the entire gene and 4 reported variants in our total NHW sample (n = 623). Of the 76 genotyped variants, 69 (MAF ≥5%, n = 39; MAF <5%, n = 30) passed QC and were further evaluated for their association with lipid traits. Based on the gene-based test, suggestive association was observed with apolipoprotein B (apoB) or proprotein convertase subtilisin/kexin type 9 (PCSK9) genes. Yet DNA sequencing does not identify mutations in these genes in a significant number of cases with a phenotype of ADH, suggesting that ADH has multiple genetic etiologies. METHODS: Through a combination of clinical examination, biochemical analysis, candidate gene approach and next-generation exome sequencing we investigated the genetic basis of an ADH phenotype in a proband of an Italian origin. RESULTS: We identified an in-frame three base-pair deletion in apolipoprotein E (APOE, Chromosome 19:45412053-55) resulting in a Leu167del mutation. The proband presented with an acute myocardial infarction at age 43, requiring urgent coronary revascularization. He had extensive tendinous xanthomas, xanthelasmata and elevated levels of total cholesterol 11.2 mmol/L, LDL-C 9.69 mmol/L, normal HDL-C 1.62 mmol/L and triglycerides levels 1.13 mmol/L. HPLC lipid profile showed selective increase in LDL-C. DNA sequencing did not identify any rare variants of the LDLR, PCSK9, LDLR adapter protein-1 (LDLRAP1) and exon 26 of the APOB gene. We then performed exome sequencing on three individuals from the family. Using data derived from exome chip genotypes for association with LDL-C, the strongest evidence of association was found for the identified APOE Leu167del mutation. CONCLUSIONS: The Leu167del mutation is predicted to alter the transmembrane structure of ApoE near the helix within the receptor binding domain. This report confirms that ADH can be caused by mutations within the APOE gene and represents the 4th loci causing ADH.
2210F Role of Titin gene variants in human dilated cardiomyopathy. R.L. Begay1, S.L. Graw1, G. Sinagra2, M. Merlo3, D. Slavov4, G. Barbati5, A. Di Lenarda6, Y. Zhu7, L. Mestrioni1, M. Taylor1. 1) Department of Medicine, University of Colorado - CU Cardiovascular Institute, Aurora, CO; 2) Cardiovascular Department, University of Trieste, Italy.

Background: The Titin gene (TTN) is the second largest human gene and encodes the largest known human protein, titin, that plays a central role in sarcomere organization. Mutations in TTN have been difficult to study due to the large size of the gene, however recent data have shown that TTN truncation mutations lead to a dilated cardiomyopathy phenotype. The large size of TTN means that almost all individuals carry missense TTN variants, the majority of which are presumably benign yet raise challenges in the evaluation of TTN in patients with overt dilated cardiomyopathy. Methods: Our group sequenced 313 exons covering the N2B and N2BA cardiac isoforms of TTN in 135 subjects with dilated cardiomyopathy. Bioinformatic filtering for ‘severe’ missense variants was based on SIFT score, GERP, Polyphen2 HDVAR, low frequencies in the 1000 genome project, and Exome Sequencing Project. Results and Conclusions: Ultimately, 45 TTN ‘severe’ missense variants were noted in 38 probands. Segregation of ‘severe’ TTN missense variants was demonstrated in larger families. Seven subjects were compound TTN heterozygotes, and six subjects were double heterozygotes with a ‘severe’ TTN variant and a pathogenic variant in another dilated cardiomyopathy gene. Genotype-Phenotype analyses between TTN ‘severe’ variant carriers and non-carriers was also performed which showed a trend towards better survival in TTN ‘severe’ missense carriers (p=0.071). These data provide important insight for researchers and laboratory into the prevalence and phenotypic consequences of ‘severe’ TTN variants in dilated cardiomyopathy patients.

2211W Desmosomal and titin gene variants in arrhythmogenic right ventricular cardiomyopathy: genotype-phenotype correlations. F. Brun1,2, C. Barnes1, G. Sinagra1, D. Slavov4, G. Barbati5, Y. Zhu7, R. Begay1, S. Graw1, B. Pinamonti1, E. Salcedo1, M. Taylor1, L. Mestrioni1, Familial Cardiomyopathy Registry. 1) Cardiovascular Institute, University of Colorado, Aurora, CO; 2) Department of Cardiology, Hospital and University of Trieste, Trieste, Italy.

Background Arrhythmogenic right ventricular cardiomyopathy (ARVC) is caused by abnormalities in desmosomal proteins of the intercalated disc and desmosomal gene mutations are the principal cause of ARVC. Recently novel variants were discovered in the sarcomeric gene titin (TTN) that are associated with ARVC. Whether known mutation carriers differ clinically from non-carriers is not well known. To address this question, we analyzed clinical outcomes in our ARVC population based on mutation status.

Methods Thirty-eight ARVC families (66 patients) were analyzed, with a median follow-up of 77 months. Genotype-phenotype association analysis was performed, and multiple variables including symptoms, electrocardiogram/echocardiogram abnormalities, arrhythmias, pacemaker and/or ICD implantation and survival free from death or heart transplant were compared between desmosomal mutation carriers, TTN carriers and non-carriers.

Results Seven patients (11%) harbored rare genetic variants in desmosomal genes (DSP, PKP2, DSG2, and DSC2), 14 (21%) carried TTN variants and 45 (68%) were non-carriers. Desmosomal carriers (DC) had a higher prevalence of inverted T waves in V2-3 in the absence of RBBB (100% vs. 32%, p<0.001) and epsilon waves (57% vs. 14%, p<0.007) compared to non-carriers. The TTN group had significantly more supraventricular arrhythmias (atrial fibrillation, atrial tachycardia) (43% vs. 0%, p=0.04) and required more pacemakers (57% vs. 0%, p=0.018). Conversely, DC required more heart transplants relative to non-carriers (57% vs. 11%, p=0.03) and exhibited a worse survival free from death or heart transplant (63% vs. 88% at 30 years and 42% vs. 88% at 50 years, p<0.001).

Conclusions This study provides valuable insights into the clinical consequences of gene mutations in individuals with ARVC. TTN rare variants confer greater risk for supraventricular arrhythmias and the need for pacemaker implantation relative to DC, while DC portends a greater risk for electrocardiogram abnormalities and the combined end-point of heart transplant or death compared to non-carriers.

2212T Burden of rare variants in PON1 is associated with ischemic stroke. D.S. Kim1,2, D.R. Crosslin1,2, P.L. Auer2,4, A.A. Burt5, A.S. Gordon4, C.E. Furlong1,2, J.F. Meschia5, M. Niwa1, U. Peters6, S.S. Rich7, D.A. Nickerson2, G.P. Jarvik1,2, NHLBI Exome Sequencing Project (ESP). 1) Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle, WA; 2) Department of Genome Sciences, University of Washington, School of Medicine, Seattle, WA; 3) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 4) Zilber School of Public Health, University of Wisconsin-Madison, Milwaukee, WI; 5) Department of Neurology, Mayo Clinic, Jacksonville, FL; 6) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD; 7) Center for Public Health Genomics, University of Virginia, Charlottesville, VA.

Background: HDL-associated paraoxonase 1 (encoded by PON1) is an enzyme with broad substrate specificity and whose activity has been associated with numerous human diseases, including ischemic stroke. Common variants in PON1 have not consistently been associated with stroke. Rare coding variation that is predicted to alter PON1 enzyme function may be more strongly associated with ischemic stroke. Methods and Results: The NHLBI Exome Sequencing Project (ESP) sequenced the coding regions of the genome (exome) in 6,503 participants ascertained for heart, lung and blood phenotypes. In this sample of 4,204 participants, 496 had verified ischemic stroke. A total of 29 nonsynonymous variants were identified in PON1. Sequence kernel association testing (SKAT) analysis adjusting for covariates identified evidence for association (p=1.29×10^-3; permutation p=3.01×10^-4) with stroke. The association of PON1 with ischemic stroke was stronger in participants with African ancestry (AA p=5.73×10^-5; permutation p=1.18×10^-5). Ethnic differences in the association between PON1 with ischemic stroke could be due, in part, to the effects of PON1 variants on missense variant (overall p=7.88×10^-5; AA p=6.52×10^-4), found at higher frequency in AA participants (1.16%) than to European ancestry (0.04%) participants. Conclusions: Rare nonsynonymous exonic variation in PON1 was associated with ischemic stroke, with stronger associations identified in those of African ancestry. The high morbidity and mortality of stroke and the higher prevalence of stroke in individuals with African Ancestry, further studies would replicate these findings and also functionally validate the effects of the rare nonsynonymous variants on PON1 enzyme function.

2213F The spectrum and prevalence of genetic background noise in patients with arrhythmogenic (right ventricular) cardiomyopathy. T.T. Kooiman1,2, S. Walker2, G. Kaur3, S.W. Scherer1,2, C.R. Marshall1,2, R.M. Hamilton1,2. 1) Physiology and Experimental Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 2) The Centre for Applied Genomics and Genetics and Genome Biology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 3) McLaughlin Centre and Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 4) Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada.

Introduction: Arrhythmogenic (right ventricular) cardiomyopathy (ACM, previously known as ARVC) is histopathologically characterized by progressive fibrofatty replacement of myocardium, primarily of the right ventricle. Disease-associated alterations in ACM are associated with the disorder have been identified in genes mostly encoding proteins that form desmosomes; however, in about 50% of patients no mutation is found in the known ACM associated genes and the pathogenicity of many identified variants is unclear. To determine the spectrum and prevalence of genetic background noise (GBN) we performed genetic variation analysis in 19 ACM cases and 35 controls and performed mutation burden analysis.

Methods: In this study, we performed whole exome sequencing of 19 unrelated ACM patients from European descent and 35 unrelated ethnicity matched control individuals and focused on the contribution of genetic variation in 52 genes previously associated with ACM, or interaction partners of these genes, to ACM susceptibility. Results: The mean number of variants identified in the 52 candidate genes in ACM patients was 53.8 per individual; controls had an average of 41.7 variants (P = NS). Mutation burden (defined as the mean number of missense, nonsense, frameshift and other deleterious mutations) of the identified variants revealed that non-synonymous rare variants (MAF<1%) in the 52 genes were not significantly more present in patients compared to controls (mean = 2.47 in cases versus 2.34 in controls). Non-synonymous rare variants in the previously ACM associated genes (PKP2, JUP, RYR2, PLN, TMEM43, TTN) also showed no significant difference in mutation burden (mean = 1.84 in cases versus 1.54 in controls). Interestingly, the overall yield of variants in the ACM genes was 73.7% in cases versus 80% in controls. Of these, 71% (cases) and 57.1% (controls) carried >1 mutation.

Discussion: This study is an evaluation of genetic variation in ACM patients versus healthy controls for the ACM susceptibility and candidate gene. Further research will be required to elucidate the role of the identified variants in ACM.

Abdominal aortic aneurysms (AAA) are a major cause of morbidity and mortality in the elderly and are characterized by extracellular matrix (ECM) degeneration, dysregulation of vascular smooth muscle cells (SMC) cells, inflammation and atherosclerosis. In approximately 20% of cases abdominal aneurysms are familial (AAA). No genes involved in AAA have been identified so far. The aim of the current study was to investigate if copy number copy number variants (CNV) may add in discovering pathophysiologic pathways and genes involved in AAA formation. Methods: AAA patients referred for genetic counseling in 2011 and 2012 were included. Patients were classified as familial when at least one first-degree was reported to have an aortic aneurysm. Patients without a positive family history were classified as sporadic (spAAA). Microarray was performed in all index patients (i.e. first relative diagnosed with AAA). Microarray was performed by using the Illumina HumanCytoSNP-12v1.1 and analyzed using Nexus CopyNumber Discovery v6.1 software. Results were evaluated using the UCSC Genome Browser March 2006, the Database of Genomic Variants and an in-house database of controls. Rare CNV regions larger than 5Mb were classified as variants and CNV regions larger than 3Mb that have been observed previously in control populations at least three times were classified as polymorphisms. GeneNetworks and Ingenuity pathway (IPA) analysis was performed of the genes identified in CNV variants. Results: The study population consisted of 67 index AAA patients; 54 were classified as fAAA and 13 as spAAA. We found 12 rare CNV variants in 11 (20%) of the fAAA patients. In patients with spAAA no rare CNV were found. Tissue expression, biological process and KEGG of the genes identified in rare CNV in fAAA was performed showing frequent involvement in inflammation, glucose metabolism, smooth muscle cell or cartilage functioning. Expression of a number of genes in cicatrization may indicate links with collagen pathways. Conclusion: These findings together with the IPA showing that a number of CNV genes are involved in AAA mechanisms may be activated as in the known genetic TGF-β, collagen, or smooth muscle cell related genetic aneurysm syndromes.

Whole-Genome Sequencing of Families with Early-Onset Myocardial Infarction. A. Mehta1, N. Schnetz-Boutaud2, D.J. Van Booven1, L. Wang3, Z. Liu1, N. Vasudeva2, M.A. Pericak-Vance1, J.L. Haines1, G.W. Beecham1. 1) University of Miami, Institute for Human Genomics; 2) Vanderbilt University, Center for Human Genetics Research. Coronary heart disease is the leading cause of death in the United States, with myocardial infarction (MI) being a primary clinical presentation. While genome-wide association studies of common variation have revealed a number of risk loci for MI, most of the loci have very small risk effects (odds ratios between 1.06-1.20) and they collectively explain less than 10% of the heritability of MI. To investigate the role of rarer variation in MI, we are performing whole-genome sequencing on 28 affected individuals with MI across 6 different families. Each family has at four or more individuals affected by MI, with at least one affected individual with onset under age 55. Whole-genome sequencing of these individuals will be performed using the Illumina HiSeq 2000 sequencer. Each individual will be sequenced using four lanes of the flowcell achieving a depth of coverage across the genome. The alignment will be conducted using the GATK pipeline, and variant and indel calling will be done using the Unified Genotype caller from the GATK pipeline. The variants and indels will be annotated using VariantServer for common annotations (e.g., cardiovascular related loci, ENCODE elements, and enhancers, etc). Together, these tools provide comprehensive annotations, like functional group, conservation scores and functional prediction scores (Sift and PolyPhen2), frequency data from the 1,000 Genomes Project and the NHLBI Exome Sequencing Project. Additional families that do not have WGS data will be analyzed using a genome-wide genotyping platform (i.e. exome chip) to establish phase and haplotype sharing within families, using the MERLIN software. Shared variants will be filtered based on annotation to identify those that are copy number variants. The most high-priority variants will be assessed in an independent dataset.

Exome sequencing identifies a novel candidate gene, NRG1, for serum cholesterol levels in Mexican families. E. Nikkola1, M. Alvarez2, M.V.P. Ling Reddy3, A. Ko1, D. Weissglas-Volkley1, C. Gutierrez-Cirlos3, L. Ribas2, M.L. Ordonez-Sanchez1, Y. Segura Kato2, T. Tusie-Luna2, C. Aguilar-Salinas2, P. Pajukanta1. 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles; 2) Instituto Nacional de Ciencias Medicas y Nutricion, SaludPublica, Mexico City; 3) Instituto de Investigaciones Biomedicas de la UNAM, Mexico City, Mexico.

Background: It is important to identify and functionally characterize genes decreasing serum low-density lipoprotein cholesterol (LDL-C) levels to better prevent atherosclerotic disease, one of the most important risk factors of coronary heart disease. The Mexican population is especially prone to hypercholesterolemia with 43% of Mexicans exhibiting total cholesterol levels >200mg/dl in a recent national survey. Familial hyperbetalipoproteinemia (FBL) is an autosomal dominant condition defined by low levels of plasma LDL-C, total cholesterol (TC), and apolipoprotein B (apoB). FHL has previously been shown to be caused by mutations in the APOB, PCSK9, and ANGPTL3 genes, and loci on chromosomes 10 and 3p21 have also been linked to FHL. However, in most FHL cases the underlying cause is unknown. To identify genes involved in FHL and cholesterol metabolism in Mexicans, we explored two Mexican FHL families without known mutations in the previously identified genes. The probands of the families had a similar kinetic profile, with an increased apoB-LDL and apoB-VLDL catabolic rate.

Methods and Results: The affected and unaffected individuals from the Mexican FHL families were exome sequenced by capturing 62M coding and regulatory regions. We focused on the variants shared by all affected family members and not present in the family members with no clinical symptoms, and restricted their type to those with high frequency, gene expression, and functional predictions using SIFT and PolyPhen. In family 1, filtering for novel missense variants and variants with minor allele frequency (MAF) <5% as well as applying the filters for gene expression and reduced the variant pool to nonsynonymous variants. In family 2, the same filtering resulted in 2 nonsynonymous variants. To further investigate whether any of these 2 genes regulates cholesterol levels generally in Mexicans, we tested their common variants in association studies in 3,700 Mexican individuals. Variants in one gene, neuregulin 1 (NRG1), passed the Bonferroni correction for association with TC, with rs13383966 providing the strongest signal. Conclusions: Exome sequencing of FHL families and a follow up in 3,700 individuals identified NRG1 gene for serum cholesterol levels in Mexicans. Replication in additional Mexican cohorts and functional studies are warranted to further investigate the role of NRG1 in cholesterol metabolism.
2218T Survival is determined by mutation type and molecular mechanism in Ehlers Danlos Syndrome (EDS) type IV. M.G. Pepin1, U. Schwarz2, D. Leistritz2, M. Lu3, P.H. Byers4, 1) Dept Pathology, Univ Washington, Seattle, WA; 2) Dept Biostatistics, Univ Washington, Seattle, WA.

EDS type IV, the vascular type, is caused by dominant mutations in COL3A1. AFFECTED individuals experience life-threatening arterial dissection and rupture, bowel perforation, and uterine rupture. We present a summary of natural history and genotype-phenotype relationships for 1232 individuals (632 index cases and 600 relatives) with EDS type IV. Median survival was 51 years. Arterial rupture and dissection were the leading cause of first complications (70%) and the leading cause of death (90%). Survival differed significantly by type of mutation (haploinsufficiency, substitutions for glycine residues in the triple helix, and splice site mutations) and even the mildest genotype (null mutations) had lower survival than the US normal. Hazard ratios for donor site mutations (8,8) and any substitution for a triple helical glycine (4,6) compared to null mutations were significant. Among substitutions for glycine in the triple helical domain, substitutions that introduced serine had a 2.6 fold higher hazard than null mutations, while for substitutions by valine the hazard was 7.2 fold higher. The location of substitutions for glycine residues in the triple helical domain was not a significant factor in survival. The median survival for females was significantly greater (54 yrs) than for males (46 yrs). Pregnancy did not alter overall survival risk (Murray et al in press). Aortic rupture in young males was one factor that contributed to the poorer survival in males. Both intracranial aneurysm and coronary artery dissection presenting as "heart attack", particularly in women in the fourth decade, were more prevalent among those affected than previously appreciated. Emergency surgical repairs of arterial rupture or dissection were often unsuccessful, while planned surgery of an identified aneurysm had significantly better outcome. In instances when a death occurred during surgical repair for colon rupture (15% of recorded cases), unexpected vascular rupture was often noted. All data suggests that surveillance of the arterial tree by MRI from the time of diagnosis may be beneficial. The findings provide some guidance for surveillance and indicate that in the creation of clinical trials attention has to be paid to mutation type to balance treatment and control groups. (Supported in part by funds from the Freudmann Fund at the University of Washington.)

2219F GENOMIC ANALYSIS OF LONG QT SYNDROME IN INDIAN COHORT. S.F. Qureshi1, C. Narasimhan2, A. Ali3, A. Venkateshwar1, K. Thangaraj4, P. Nallari1, 1) Department of Genetics, Osmania University, Hyderabad, Andhra Pradesh, India; 2) Care Hospital, Hyderabad; 3) Institute of Genetics and Hospital for Genetic Diseases, Osmania University, Hyderabad; 4) Centre for Cellular and Molecular Biology, Hyderabad.

Long QT Syndrome, characterized by prolonged QTc along with congenital bilateral deafness is a result of homozygous or genetic compound variations in either of the K+ channel genes, KCNQ1 and KCNE1. Tumor Necrosis factor-alpha, a known inflammatory marker is one of the modifier genes in CVDs. Here we report results on analysis of KCNQ1 and KCNE1 transcripts in 50M; 50F, 32 LQTS patients and the available family members followed by In-silico analyses to establish the genotype-phenotype correlations. Genomic and Exonic variations were further confirmed by automated DNA sequencing. Results: The present study revealed 3 novel intronic and 2 novel synonymous variations and 2 cases of compound heterozygosity, apart from 1 reported missense variation in KCNQ1 and 1 reported a silent mutation in KCNE1 gene. Screening of MYH7 revealed 3 novel missense variations and 1 novel synonymous variation which will be discussed. Although these variations are not highly pathogenic but might have a regulatory role in the splicing mechanism, a role their effect might be influenced in the presence of modifier genes and environmental factors.

2220W Novel sarcomeric gene variations in Hypertrophic cardiomyopathy patients of Indian Cohort. A. Rangaraju1, M.L. Satyanarayana2, C. Narasimhan1, P. Nallari1, 1) Department of Genetics, Osmania University, Hyderabad, Andhra Pradesh, India; 2) Cardiologist, CARE Hospitals, Hyderabad, Andhra Pradesh, India.

Hypertrophic Cardiomyopathy (HCM) is a major cause of morbidity and mortality, characterized by hypertrophy of the left/right ventricle and intervening ventricular septum with an autosomal dominant mode of inheritance. It has a prevalence of 1 in 500 and exhibits a variable clinical course ranging from being asymptomatic to severe heart failure and sudden cardiac death. Mutations in two sarcomeric genes that encode myosin heavy chain and their supporting proteins have been identified as a cause of HCM. Despite the identification of 900 different causative mutations, the frequency of mutation rate is very low, alternatively, the effect of modifier genes could influence the degree of hypertrophy developed in response to the causal mutation. MYBPC3 and MYH7 are the two genes which have been largely implicated accounting to 50% of HCM cases. Objective: Considering the phenotypic and genotypic heterogeneity of HCM and the ethnic/geographic variation existing in the Indian population, screening for the possible genetic variations in MYBPC3 and MYH7 in HCM patients of Indian cohort was carried out. Methodology: Blood samples of 100 controls and 100 patient samples were collected after obtaining the informed written consent along with the clinical data of the patients for genomic DNA isolation. For screening MYBPC3 and MYH7 genes, PCR based SSCP analysis was carried out. The samples with variation were further confirmed by automated DNA sequencing. Results: The present study revealed 3 novel intronic and 2 novel synonymous variations and 2 cases of compound heterozygosity, apart from 1 reported missense variation in KCNQ1 and 1 reported a silent mutation in KCNE1 gene. Screening of MYH7 revealed 3 novel missense variations and 1 novel synonymous variation which will be discussed. Although these variations are not highly pathogenic but might have a regulatory role in the splicing and translation mechanism, a role their effect might be influenced in the presence of modifier genes and environmental factors.

2221T A family-based whole genome sequencing study to identify genetic determinants of platelet hyper-aggregation following aspirin in African Americans at high risk for coronary artery disease. M.A. Taub1, L.R. Yanek1,2, I. Ruzek1, L. Huang2, D.M. Becker3, L.C. Becker3, R.A. Mathias1,2, 1) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Johns Hopkins School of Medicine, Baltimore, MD; 3) The GeneSTAR Program, Johns Hopkins School of Medicine, Baltimore, MD.

Background: Low dose aspirin (ASA) is given to inhibit platelet aggregation in individuals at risk for arterial thromboses and is considered standard care for secondary prevention of coronary artery disease (CAD). We have previously identified common genetic determinants of platelet aggregation following 14-days of ASA, 81 mg/day, using a GWAS approach, and here build a long-standing study of familial aggregation for CAD (GeneSTAR) applying a whole genome sequencing (WGS) approach. Methods: 5 African American families segregating the phenotypes of hyper-aggregators (>90th percentile of trait distribution in the full sample, N=869) and normo-aggregators (<50th percentile) to collagen (5 uM) post-ASA, 30 subjects underwent WGS by Illumina, Inc. (depth >30X). Variants were filtered (GO>20 and segmental duplication regions masked). We determined the set of variants showing perfect co-segregation with phenotype in each family. We tabulated the number of families with perfect co-segregation at (i) each variant site; and (ii) variants collapsed over the units of CCDS transcripts focusing on variants novel to GeneSTAR, i.e. not observed in The Thousand Genomes Project. Results: We observed −4 million (M) variants per genome, of which ~13K were private variants not shared with another family member. Of the 12.7M sites present across all the samples, 1.6M were novel to GeneSTAR. While we found no sites that showed perfect co-segregation with phenotype in all 5 families, variants in the promoters of CDK1 and SLCE1A1 co-segregated with phenotype in 3 families. No rare variants were found in common across all 5 families. In conclusion, gene ontology (GO) analysis of the 218 transcripts that co-segregated with phenotype across 4/5 families includes cell adhesion and wound healing as two of the most significant hits (hypermorphic p-value<0.01). Restriction to only long-studied GO categories revealed~140 for GO cell adhesion, and for most families with a GO enrichment in epithelial cell-cell adhesion and cell adhesion genes (p-value<0.01). Conclusions: This analysis suggests a potential role for rare sequence-identified variants in pathways of genes as determinants of platelet hyper-aggregation post ASA. Analysis is underway to assess statistical significance for these variants, extend annotation beyond coding sequence by leveraging ENCODE data, and test for replication in a sample of 100 independent WGS African American samples from GeneSTAR.
2222F
The molecular dissection of familial dilated cardiomyopathy, J.M. Taylor, S.L. Reid, J. Hayesmoore, J. Woodley, K. McGuire, K. Thomson, E. Blair, W. Watkins, A. Seller, 1) Oxford Medical Genetics Laboratory, Oxford University Hospitals NHS Trust, Oxford, Oxfordshire, United Kingdom; 2) Department of Clinical Genetics, Oxford University Hospitals NHS Trust, Oxford, Oxfordshire, United Kingdom; 3) Department of Cardiovascular Medicine, West Wing, John Radcliffe Hospital, Oxford OX3 9DU.

Introduction: Familial dilated cardiomyopathy (DCM) exhibits considerable clinical heterogeneity both within and between families: penetrance is incomplete and age dependent. Combined clinical screening and molecular analysis is therefore essential to identify affected individuals. Currently, there are in excess of 30 genes reported to cause nonsyndromic familial DCM. This study focuses on the genetic heterogeneity of 50 familial DCM patients from a single referral centre and aims to identify the frequency and distribution of clinically relevant variants within this discrete patient cohort.

Methods: Molecular analysis of 28 DCM-related genes was undertaken within the Oxford Genetics Laboratory. Enrichment of targeted genes was undertaken using a custom built Haloplex™ kit (Agilent) and sequenced using an Illumina MiSeq. A commercial software package (NextGEne™, by Softgenetics) was used for data alignment, which was then filtered using an in-house pipeline.

Results: On average, 97.5% of targeted regions were consistently covered to ≥30 read depth and clinically relevant variants were detected in ~40% of patients. The test sensitivity is higher than expected (previously reported as 17.35-17.2%). Pathogenic variants were found within genes responsible for the sarcomere and desmosome structure, and the sodium ion channel.

Conclusion: This study demonstrates the utility of an extensive DCM gene panel and further characterizes the genetic heterogeneity. Clinically relevant novel genotype-phenotype associations will be discussed.


2223W
Novel homozygous missense mutations in the SLC2A10 gene in a Turkish pediatric patient with arterial tortuosity syndrome. S.G. Temel, O. Bostan, D. Proost, L. Van Laer, E. Cel, B. Loeyes, 1) Histology & Embryology, University of Uludag, Bursa, Turkey; 2) Department of Paediatric Cardiology, University of Uludag, Faculty of Medicine, Bursa, Turkey; 3) Center for Medical Genetics, Antwerp University Hospital and University of Antwerp, Antwerp, Belgium.

Arterial tortuosity syndrome (ATS) is a rare autosomal recessive connective tissue disorder, mainly characterized by tortuosity and elongation of the large- and medium-sized arteries with predisposition to stenoses and aneurysms. ATS is caused by mutations in the SLC2A10 gene, encoding for the facilitative glucose transporter 10 (GLUT10) and is described typically for the sarcomere and desmosome structure, and the sodium ion channel. Pathogenic variants were found within genes responsible for the sarcomere and desmosome structure, and the sodium ion channel.

2224T
Mutation detection in aortopathy and other vasculoopathies complicating hereditary disorders of connective tissue by next generation sequencing. E. Zaklyazminskaya, R.A. Weerakkody, J. Biggs, P.J. Norris, C. Neuworth, L. Game, A. Vandersteen, F.M. Pope, N.J. Cheshire, T.J. Altmann, 1) Physiological Genomics and Medicine Group and Genomics Laboratory, MRC Clinical Sciences Centre, Imperial College London, UK; 2) National Ehlers-Danlos Syndrome Diagnostic and Treatment Regional Genetics Service, North West London Hospital NHS Trust, UK; 3) Imperial Vascular Unit, St Mary’s Hospital, Imperial College, London, UK.

Hereditary disorders of connective tissue are a heterogeneous group of disorders with overlapping clinical features. A proportion of these patients are at risk of early arterial rupture. Early diagnosis, screening of family members and subsequent antihypertensive therapy and vascular surveillance can therefore be life-saving. Some patients with connective tissue disorders have clinical features that suggest a specific diagnosis such as Marfan syndrome, or vascular Ehlers-Danlos syndromes (EDS), but most patients do not clearly fit such a diagnosis. Genetic testing is available but due to the cost and time associated with Sanger sequencing these tests are usually limited to familial and syndromic forms of the disease. DNA samples were obtained from 300 patients with thoracic aortic aneurysms and dissections (TAAD) and/or with a vascular EDS phenotype. To test the utility of next generation sequencing for detecting mutations in these disorders, amplicons were designed from exons and exon-intron boundaries of genes known to be associated with TAAD (FBN1, TGFBR1, TGFBR2, MYH11, ACTA2, MYLK) and EDS (COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, TNXB). Samples were amplified using the Access Array System (Fluidigm) and sequenced by MiSeq (Illumina). In the pilot study of 126 samples, 96.1% of bases were covered at 25X and 89.1% were covered at 50X. Rare likely pathogenic variants were identified in 15% of aortopathy patients (n=70) and 10% of vascular EDS patients (n=56). Analyses on a further 174 patients are ongoing. Our data indicate that pathogenic mutations in genes previously thought to be associated with rare familial or syndromic forms of aortopathy may be implicated in a significant proportion of sporadic cases and demonstrate the potential utility of next generation sequencing for routine diagnostics of hereditary disorders of connective tissue.

2225F
New frame-shift deletion in EMD gene causes Emery-Dreifuss muscular dystrophy with severe cardiomyopathy required heart transplantation. E. Zaklyazminskaya, O. Blagova, M. Yakovleva, V. Pumiyanteva, A. Nedostup, D. Shumakov, S. Dzemeshkevich, 1) Medical Genetics Laboratory, Petrovsky Moscow Russian Research Centre of Surgery RAMS, Moscow, Russian Federation; 2) Sechenov Moscow State Medical University, Moscow, Russian Federation; 3) Federal Research Centre of Transplantation and Artificial Organs, Moscow, Russian Federation.

Background: Emery-Dreifuss muscular dystrophy (EDMD) is a clinically heterogeneous disease characterized by progressive muscular atrophy, joint contractures, normal intelligence, and variable cardiac involvement. At least 6 genes are known to cause EDMD when mutated (EMD, LMNA, SYNE1, SYNE2, FHL1, and TMEM43). Clinical case. Male proband, 38 y.o., has addressed to genetic counseling because of low progressive skeletal myopathy from the childhood, having walking difficulties but ambulant, moderate knees and elbows contractures, high CK level, normal intelligence, arrhythmias (HR 83-96 bpm, paroxysmal atrial fibrillation, AVB II, RBBB, premature ventricular beats), and dilated cardiomyopathy (EF LV 33%). Proband’s mother died at the age of 50 due to clinical signs of EDMD at the age of 53 y.o. Two sons, 4 y.o. and 11 y.o. are healthy. Informed consent for genetic testing was taken in accordance with Helsinki declaration. Mutations analysis was performed by PCR-based direct Sanger sequencing of coding area and adjacent intronic areas of genes of interest. Two genetic variants were detected: frame-shift deletion c.del619C in EMD gene and c.IVS4-13T>A in LMNA gene. Both variants were not found in control group of healthy donors.

Taking into account progressing of cardiac arrhythmias and high risk of sudden cardiac death in patients with EDMD, dual-chamber ICD was implanted. Patient had repeated appropriate shocks 6 month after ICD implantation despite anti-arrhythmic therapy, and progressive impairment of cardiac pump function. Heart transplantation was proposed and successfully performed. There was no anesthetic complication during the procedure. For now, the follow-up period is 3 month with significant health improvement. Conclusion. We detected a new frame-shift mutation c.del619C in EMD gene causing premature stop-codon appearance and protein shortening (p.E236X). We suggest pathogenic role of this mutation in family DCM of EMD with cardiac manifestations. The role and possible modulating effect of rare novel VUS c.IVS4-13T>A in LMNA gene is to be elucidated. Study of possible effect on LMNA mRNA processing is in progress now.
Assessment of the enrichment for rare coding variants in 16 related cases of fibromuscular dysplasia. N. Boualla-Najj1,2, R. Kiando1,2, P.F. Pilot1,2,1, X. Kromer1,2,1. Paris Cardiovascular Research Center, U970 HEGP research Center, INSERM, Paris, France; 2) Paris Descartes University, Paris Cité Sorbonne, Paris France; 3) AP-HP, Centre for Rare Vascular Diseases, Hopital Européen Georges Pompidou, Paris, France.

Fibromuscular dysplasia (FMD) is an arterial disease characterized by non-atherosclerotic stenoses reported mostly in renal (70%) and extra-cranial carotid (30%) arteries. FMD predisposes to hypertension, renal ischemia and stroke, the first cause of disability worldwide. The causes of FMD are unknown, but predominant manifestations have a genetic component years with a prevalence of ~4/1000. The genetics of FMD is under-investigated because of the lack of large families and cohorts due to the rarity of the disease and the complexity of the diagnosis that requires artery biopsy and/ or complex imaging. The aim of this study was to assess the role of rare coding and putatively functional variants in the onset of FMD in related cases. Patients are five sib-pairs and two sib-trios, all affected (N=16) with confirmed multifocal string-of-beads FMD diagnosed by angiography/scaner. Mean age diagnosis is 4/4yrs (min=23; max=57). Exome sequencing and variant calling was performed by Integragen® using Agilent’s capture Kit (V4) and HiSeq2000 (Illumina®) with 5×4 median coverage. We identified 4,311 confident (high calling-score and depth >8X) missense rare (MAF<0.01 in EVS and/or 1000G) variants predicted to be damaging (Polyphen2) in 3,414 genes. No gene carried mutations that were specific to all patients or at least 3 out of 7 sibpairs/trios suggesting strong genetic heterogeneity in familial FMD. In order to identify genes to prioritize that might be enriched in mutations in FMD cases, we performed a genomic burden test with highly mutated genic regions. Not only a few genes from 4,300 unselected Europeans as controls (Exome Variant Server).

Fischer exact test that compared variants vs non-variant alleles counts in cases and controls revealed several genes significantly (P<10e-5) enriched for putatively functional variants at the genomic level. Important fibrosis and loss of extracellular matrix (ECM) at the media of FMD arteries has been demonstrated. Using less significant threshold (P<0.01), we identified two genes involved in ECM structure, cell shape and junction, which are interesting candidates for FMD. Validation of the enrichment of statistical significant genes requires the extension of the current sample to larger cohorts of patients either by exome sequencing or genotyping.

Resequencing of Renin-Angiotensin-Aldosterone-System Genes Identifies Rare Variants Associated with Blood Pressure Salt-Sensitivity: The GenSalt Study. T.N. Kelly1, J.E. Hixson2, L.C. Shimmin2, Q. Zhao2, D. Gu2, J. He1. 1) Dept Epidemiology, Tulane Univ, New Orleans, LA; 2) Human Genetics Center, University of Texas, Houston, TX; 3) Cardiovascular Institute and Fujiw University Medical Center, Chinese Academy of Medical Sciences and Peking Union Medical College, and Chinese National Center for Cardiovascular Disease Control and Research, Beijing, China.

Genetic association studies have identified significant associations between common variants from renin-angiotensin-aldosterone system (RAAS) genes and blood pressure (BP) responses to dietary sodium intervention. However, the role of rare RAAS variants in salt-sensitivity remains unexplored. We resequenced a resequencing rare functional and coding variants associated with BP salt-sensitivity among participants of the Genetic Epidemiology Network of Salt-Sensitivity (GenSalt) study. The GenSalt study was conducted among 1,906 participants who underwent a 7-day low-sodium (51.3 mmol sodium/day) followed by a 7-day high-sodium feeding study (307.8 mmol sodium/day). We selected 300 GenSalt subjects with the highest and 300 GenSalt subjects with the lowest mean arterial pressure responses to the high sodium intervention to participate in the current resequencing study. Functional regions of seven RAAS genes, including ACE2, APLN, AGTR1, HSD11B1, HSD11B2, NR3C2, and RENBP were resequenced using the VariantSEQTM system (Applied Biosystems; Foster City, CA). RAAS variants with minor allele frequencies less than 5% were collapsed according to gene and analyzed using the cohort allelic sums test (CAST). We identified significant associations over coding variants of the APLN, AGTR1, and HSD11B2 genes and BP salt-sensitivity, with p-values of 0.05, 0.03, and 0.03, respectively. Within the promoters, splice sites, exons, and 3’ untranslated regions of these 3 genes, we identified 17 rare APLN variants in 10 rare AGTR1 variants, and 19 rare HSD11B2 variants. Nine percent of GenSalt participants with high salt-sensitivity were carriers of at least one of the rare APLN variants, while only 4% of salt-resistant participants were carriers. In addition, approximately 17% of participants with high salt-sensitivity carried 11 rare AGTR1 variants, and only 11% of salt-resistant subjects were carriers. Further, 8% of those who were highly salt-sensitive compared to only 4% of those who were salt-resistant were carriers of HSD11B2 variants. In summary, we provide the first evidence for a role of rare and potentially functional RAAS variants in BP salt-sensitivity. Validation study will be needed to confirm these findings.
Identification of a single nucleotide polymorphism variant in TYRO3 associated with coronary artery disease risk in the ClinSeq® Study. H. Sung1, B. Suktitipat1, K. Lewis2, D. Ng2, SG. Gonsalves2, JC. Mullikin1,2, LG. Biesecker1, AF. Wilson3, NISC Comparative Sequencing Program. 1) Genomics Section, Inherited Disease Research Branch, National Human Genome Research Institute, NIH, Baltimore, MD; 2) Genetic Disease Research Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 3) Comparative Genomics Unit, Genome Technology Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 4) National Institutes of Health Intramural Sequencing Center (NIHC), National Human Genome Research Branch, NIH, Bethesda, MD.

ClinSeq® is a large-scale medical sequencing study designed to investigate associations of rare sequence variants with traits related to coronary artery disease (CAD). The study currently includes 1092 non-smoking participants between the ages of 45 to 65 with normal to severe coronary artery calcification scores. About 200 CAD-related traits were measured at the NIH Clinical Research Center in Bethesda, MD. Whole-exome sequencing was performed with the Agilent SureSelect 38Mb and 50Mb capture kits for 387 and 325 individuals, respectively, at NIH Intramural Sequencing Center. Single nucleotide polymorphism variants (SNVs) common to both capture regions with a call rate > 98% and minor allele frequency (MAF) > 1% were used to check for cryptic relatedness and for misspecified population stratification by multidimensional scaling analysis - 635 unrelated European Americans (EAs) were included. For each capture kit with EAs only, SNVs with at least one homozygote and a call rate ≥ 50% were included. The two capture regions with only SNVs in common were merged, yielding 439,807 SNVs. Of these SNVs, 74% and 46% had MAFs < 0.01 and < 0.001, respectively. Individuals were divided into four bins based on their 10-year Framingham risk scores for developing CAD: Bin 1 (< 5%), Bin 2 (5-10%), Bin 3 (> 10%) and Bin 4 (known CAD). The union of all SNVs in 232 individuals in Bin 1 were compared with the union of all SNVs in 89 individuals in Bin 3 (> 10%) and Bin 4 (known CAD). The union of all SNVs in 232 individuals in Bin 1 were compared with the union of all SNVs in 89 individuals in Bin 4 using the Variant Annotation, Analysis, and Search Tool (VAAST), a probabilistic search tool in likelihood-framework based on amino acid substitution frequencies. Under the assumption of a recessive inheritance model in terms of best-scoring homozygous or heterozygous variants in the VAAST analysis, a nonsynonymous variant TYRO3 was identified (rs62001448, MAF=0.095) with a p-value=1.01e-81. The variant allele (T) at rs62001448 was found in a heterozygous state in 16 (7%) times in 232 individuals in Bin 1 and 26 (29%) in 89 individuals in Bin 4. No variant allele (T) was found in a homozygous state. Mouse studies done by Angelillo-Scherrer at al.(2005) and Cosemans et al.(2010) indicated that a loss of function in TYRO3 would decrease thrombus formation and a gain of function would increase thrombus formation. This finding suggests the importance of TYRO3 gene in CAD and provides insight into some of the underlying genetic mechanisms of CAD in the ClinSeq® study.
Rare and Common Exome Chip Variants are Associated with Fasting Glucose and Insulin Levels - The CHARGE-S Exome Chip and Sequencing Study, 1 J. Wessel, 2 B. Cornes, 3 S. Wang, 4 R. Jensen, 5 A. Kral, 6 E. Lange, 6 M. Nalls, 7 J. Wilson, 8 B. Cade, 9 Y. Lu, 9 R. Loos, 9 V. J. Dudins, 10 J. Pankow, 11 S. Willems, 12 C. van Duijn, 12 L. Biebalak, 13 X. Guo, 14 L. Rasmussen-Torvik, 15 M. Province, 16 I. Borecki, 17 J. Dupuis, 17 J. Rotter, 17 D. Siscovick, 18 M. Goodarzi, 18 J. Maigs, 19 CHARGE Consortium.

1) Dept. of Epidemiology, Indiana Univ Sch Public Health, Indianapolis, IN; 2) Department of Medicine, Harvard School of Medicine, Boston, MA; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 4) Department of Medicine, University of Washington, Seattle, WA; 5) Department of Genetics, Washington University School of Medicine, St Louis, MO; 6) Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC; 7) Laboratory of Neurogenetics, National Institute of Aging, Bethesda, MD; 8) Department of Medicine, University of Mississippi Medical Center; 9) The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai; 10) Icelandic Heart Association; 11) Department of Epidemiology, University of Minnesota School of Public Health; 12) Department of Epidemiology, Erasmus University Medical Center; 13) Department of Epidemiology, University of Michigan School of Public Health; 14) Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA; 15) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine.

Genome-wide association studies (GWAS) have identified 56 loci in which common single nucleotide variants (SNVs) are associated with fasting glucose (FG) or fasting insulin (FI) levels. These SNVs explain $<5%$ of FG or FI variation and are largely non-exonic. To identify new and rare (MAF$<1\%$) candidate genes, we performed whole exome sequencing in 26 clinically well-characterized French-Canadian families that are followed at the Montreal Heart Institute’s genetic clinic. The inclusion criteria comprised four main conditions: patients had left ventricular ejection fraction $\leq 45\%$, left ventricular dilatation, absence of other known causes of cardiomyopathy, as well as a family history of DCM. Next-generation DNA sequencing was done using the Illumina HiSeq2000 instrument and a paired-ends 2×101 bp protocol. Results: We identified 24,714 novel variants out of which 4,863 are coding. We show that 18% of the families in our study carry unique and novel nonsense mutations in the TTN gene, consistent with previous reports. Further, we identified a nonsense mutation in BAG3, Arg309X, which co-segregated with DCM in three families. The mutation was shared by all 14 affected members and absent in all healthy individuals. Importantly, we show that BAG3 truncating mutations are significantly correlated with early onset of DCM. Conclusion: Mutations in TTN and BAG3 account for a significant proportion of familial dilated cardiomyopathy at the Montreal Heart Institute. Our results highlight the importance of whole exome sequencing in identifying novel causal mutations in DCM which could lead to a better understanding of the disease and improve clinical testing permitting earlier diagnosis in affected families.

Novel Dilated Cardiomyopathy Mutations Identified by Whole-Exome Sequencing, N. Cham1,2, R. Tadros1,2, M. Beaudoin1, K. Sin Lo1, L. Robb1, F. Lemarbre1, M. Talajic1,2, G. Letarte1,2, 1 Montreal Heart Institute, Montreal, QC, Canada; 2) Faculté de Médecine, Université de Montréal, Montréal, Québec, Canada.

Introduction: Dilated cardiomyopathy (DCM) is a disorder of the heart muscle that is characterized by a dilated and impaired contraction of the left ventricle and may ultimately lead to sudden death, heart failure or transplantation. Genetic factors play a major role in DCM; however, known mutations explain less than 20% of familial cases. We hypothesized that a whole-exome approach to identify novel mutations would improve understanding of the genetic basis of DCM. Methods: We performed whole exome DNA sequencing in 26 clinically well-characterized French-Canadian families that are followed at the Montreal Heart Institute’s genetic clinic. The inclusion criteria comprised four main conditions: patients had left ventricular ejection fraction $\leq 45\%$, left ventricular dilatation, absence of other known causes of cardiomyopathy, as well as a family history of DCM. Next-generation DNA sequencing was done using the Illumina HiSeq2000 instrument and a paired-ends 2×101 bp protocol. Results: We identified 24,714 novel variants out of which 4,863 are coding. We show that 18% of the families in our study carry unique and novel nonsense mutations in the TTN gene, consistent with previous reports. Further, we identified a nonsense mutation in BAG3, Arg309X, which co-segregated with DCM in three families. The mutation was shared by all 14 affected members and absent in all healthy individuals. Importantly, we show that BAG3 truncating mutations are significantly correlated with early onset of DCM. Conclusion: Mutations in TTN and BAG3 account for a significant proportion of familial dilated cardiomyopathy at the Montreal Heart Institute. Our results highlight the importance of whole exome sequencing in identifying novel causal mutations in DCM which could lead to a better understanding of the disease and improve clinical testing permitting earlier diagnosis in affected families.

Whole-exome sequencing of a large cohort of patients with congenital heart disease in the Quebec founder population, C. Preuss1, C. Capron1, C. Prive1, J.C. Grenier1, M. Samuels2, P. Chetelaine3, P. Awa-dalla1, G. Andelfinger1, 1) Department of Pediatrics, Centre Hospitalier Universitaire Sainte Justine, Montreal, QC, Canada; 2) Centre de Recherche de l’Hôpital Ste-Justine, Université de Montréal, Montréal, Quebec, Canada; 3) Cardiology Service, Centre Mère-Enfants, Centre Hospitalier Universitaire de Québec, Université de Laval, Quebec City, Québec, Canada.

Congenital heart disease (CHD) is the most common birth defect, the most common cause of infant morbidity and the second most common cause of infant mortality in Canada. It is defined as a structural malformation of the heart or/and great vessels that is present at birth and is of functional significance. Mendelian transmission of congenital heart disease in large families is rare, even in those for which a variable expression of penetrance is observed. We have whole-exome sequenced 96 individuals with left ventricular outflow tract obstructions such as septal defects, aortic valve lesions, coarctation, and supraventricular arrhythmias which have been recruited in Quebec. Library preparation was performed using the TargetSeq enrichment kit and sequences were obtained using the SOLID5 platform in our in house genomics facility. For several members of multiple affected families genetic linkage mapping has been performed. However, since genetic and phenotypic heterogeneity is the norm even for rare Mendelian disease such as CHD, it remains a challenging tasks to make accurate and clinically useful predictions regarding the disease phenotype using classical genetic approaches. Here, we present our recent results from whole-exome sequencing highlighting that genomic, phenotypic and population context matters. The filtering for rare, potential damaging disease variants revealed a number of loci enriched for disease related pathways including histone modification, NOTCH signaling and metalloproteins involved in heart development. Many of these rare variants identified in newly implicated candidate genes are the subject of ongoing validation and functional studies and might jointly determine the phenotype. In order to test our hypothesis whether the excess of rare, population specific variants contributed to disease risk in individuals, we will validate the identified variants and perform a population wide screening in an independent control cohort from Quebec (CARTAGENE). The heterogeneous nature and often not complete penetrance of CHD suggests that compelling effects of mutation clustering within families may be rather the norm than the exception for traits that have been primarily studied under the Mendelian mode of disease inheritance.
A comprehensive analysis study of APOE-C1-C4-C2 gene cluster variation with plasma lipoprotein traits in U.S. Whites, Z.H. Radwan1, X. Wang2, F. Wagon3, J.E. Hokanson4, R.F. Hamman5, M.M. Barmada6, F.Y. Demirio7, M.I. Kambho7. 1) Human Genetics, GSPH, Univ Pittsburgh, Pittsburgh, PA; 2) Epidemiology, Colorado School of Public Health, Univ Colorado Denver, Aurora, CO.

APOEpoliprotein (APO) are the major determinants of plasma lipoprotein-lipid distribution and their genetic variation have significant impact on cardiovascular disease (CVD) risk, which is the leading cause of morbidity and mortality worldwide. Therefore, understanding the genetic of plasma lipid traits and determining subjects with high risk of developing CVD is of great public health importance. Although the influence of APOE common polymorphisms on plasma lipid traits has been extensively evaluated in diverse ethnic groups, little attention has been paid on other genes located in the same gene cluster on chromosome 19. The objective of this study was to characterize common and rare genetic variation through comprehensive sequencing of the APOE, APOC1, APOC4, APOC2 genes along with their hepatic regulatory regions HCR-1, and HCR-2 in 95 subjects from US non-Hispanic Whites (NHWs) with extreme HDL-C/TG levels (47 in the upper 10th percentile, and 48 in the lower 10th percentile), and then to genotype the identified tagSNPs and rare variants along with HapMap tagSNPs to cover the intergenic regions in 623 NHWs to evaluate their associations with lipid traits. A total of 15 variants (105 substitution, 10 indels) were identified in the sequencing subset of which 74% were reported, 26% were novel, 54% were common (MAF≥5%), and 46% were rare or less common (MAF≤5%). The number of observed unique rare variants was higher in the high HDL-C/low TG group than in the low HDL-C/high TG group (23 vs. 15). Seventy variants contributed to 118 (59%) of the 206 candidate rare and low frequency variants in 623 subjects followed by association analyses. Gene-based association results confirm the major contribution of this gene cluster on variation in plasma LDL-C, TC, and apoB levels. Single-locus association analysis revealed 20 significant associations with at least one lipid trait (9 6.5E-13; p<0.046). Rare variants analyses showed more significant associations with TC than with other lipid traits. Our findings confirm the significant contribution of this gene cluster in affecting plasma lipid profile and consequently the CVD risk.

Whole genome sequencing in African American families to identify genetic determinants of extreme obesity. L. Yanek 1, M. Taub 1, J. Ruzzini 2, L. Huang 2, D. Valdiva 1, L. Becker 2, D. Becker 2, R. Mathias 1, The Genestar Research Program. 1) The Genestar Research Program, Johns Hopkins School of Medicine, Baltimore, MD; 2) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Johns Hopkins School of Medicine, Baltimore, MD.

Background: African Americans (AA) have the highest rate of obesity in the US. We have noted a dramatically high prevalence of extreme obesity (BMI≥35 kg/m^2) in AAs in the Genestar study of families with a history of early-onset coronary artery disease (CAD). We searched the AAs in the GeneSTAR families had >1 relative pair with extreme obesity. We leverage whole genome sequence (WGS) data available on 3 AA families containing one or more members with extreme obesity and one or more with normal BMI (18.5-25) by comparing genomic regions that segregate with extreme obesity. Methods: Deep (>30X) WGS were performed by Illumina, Inc, in 130 families with ≥10 members with extreme obesity and one or more with normal BMI by Exome Capture and targeted sequencing to the exome. We used the Hong Kong IBD algorithm to search for rare variants. We then performed a genome-wide association analysis to identify genetic variants in 623 subjects followed by association analyses. Gene-based association results confirm the major contribution of this gene cluster on variation in plasma LDL-C, TC, and apoB levels. Single-locus association analysis revealed 20 significant associations with at least one lipid trait (9 6.5E-13; p<0.046). Rare variants analyses showed more significant associations with TC than with other lipid traits. Our findings confirm the significant contribution of this gene cluster in affecting plasma lipid profile and consequently the CVD risk.

RNAseq Analysis of Congenital Heart Defects. D.M. McKean1, D.S. Herms1, A.G. Bick2, D.C. Christodoulou2, S.R. DePalma3, J.M. Gorham4, J. Hotton5, Y. Jiang6, J.D. Overton6, H. Wakimoto7, S. Zaidi8, R.E. Breibart9, W.K. Chung9, R.P. Kim3, O. Toka9, R.P. Litton9, C.E. Seidman9, J.G. Seidman9, Pediatric Cardiac Genomic Consortium. 1) Genetics, Harvard Medical School, Boston, MA; 2) Boston Children's Hospital, Boston, MA; 3) Pediatrics, Columbia University Medical Center, New York, NY; 4) Children's Hospital Los Angeles, Los Angeles, CA; 5) Institute of Human Genetics, Erlangen University, Erlangen, Germany; 6) Genetics, Yale University, New Haven, CT.

Congenital heart disease (CHD) is the most common type of birth defect and the leading cause of death within the first year of life. To elucidate genetic causes and transcriptional responses to CHD mutations, we performed next generation sequencing and RNAseq on cardiovascular tissues. From 119 CHD subjects, 160 cardiac tissues were obtained from 11 unique sites. For 23 subjects (46 cardiac samples), genomic sequence (whole exome or whole genome) and RNAseq data was available. High quality RNAseq reads were aligned to hg19 using TopHat1.4. Duplicate reads were removed, yielding an average of 20 million reads per gene. Gene expression levels were compared with tissue expression from other subjects, from comparable neonatal mouse tissues, and from different tissue from same individuals. 76 genes expressed in neonatal mouse were not expressed in human tissues and 285 genes expressed in human tissues were not expressed in neonatal mouse. 55 genes displayed allele-specific expression (ASE) in multiple human samples (perhaps reflecting cardiac imprinting). 146 genes displayed ASE in ≥1 tissue in a single subject. Interrogation of genomic DNA is under-way to consider if ASE in these subjects reflects genomic mutation. Genomic DNA was analyzed to predict RNA splice variants (Maximum Entropy, Neural Networks, Splice Site Finder, Human Splice Finder, and Gene Splicer algorithms). Over 1000 unique sites were predicted to alter RNA splicing from one or more algorithms. Among >300 genes expressed in the heart, aberrant splicing was seen with 14 variants, and not seen with 9 variants. Active studies integrating allele loss with predicted splice variants are underway. Ongoing work will define the sensitivity and specificity of splice variant prediction tools.
Novel rare coding variants underlie blood lipid levels in the population: an exome array association study in 55,000 whites and blacks.

Circulating low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG) are quantitative, heritable risk factors for coronary heart disease, and genome-wide association screens of common DNA sequence variants have identified many loci associated with these traits. Here, we test the hypothesis that rare coding and splice-site mutations contribute to inter-individual variation in blood lipid concentrations in the population. We characterized ~41,000 European Ancestry (EA) and ~14,000 African Ancestry (AA) samples from 13 studies using a new, rare-variant genotyping array ("the Exome Chip") that was designed based on the protein-coding sequences for ~18,500 genes ("the exome") in >12,000 individuals. Within each of the 13 contributing studies, we tested the association of plasma lipids with individual rare variants as well as sets of variants within each gene. To combine statistical evidence across studies, we performed fixed effects meta-analysis. For single variant association, we set the significance threshold to 7 × 10^-8, corresponding to a Bonferroni correction for 742,932 tests (3 phenotypes x 247,644 variants on the array). For the gene-based association, we set the significance threshold to 5 × 10^-7, corresponding to a Bonferroni correction for 104,455 tests (3 phenotypes x 17,574 genes on the array x 2 tests). The most significantly associated results for each trait replicated known associations in APOE for LDL-C, CETP for HDL-C, and APOA5 for TG. We identified two new genes with rare coding variants associated with lipid levels in EA samples: ANGPTL8 for HDL-C (G46X, 0.1% frequency, β = 10 mg/dL, p = 9 × 10^-11), and PAFAH1B2 for triglycerides (S161L 1.7% frequency, β = 10%, p = 4 × 10^-11). Additionally, we found two new gene associations in the AA samples: PCSK7 for HDL-C (R90H, 0.2% frequency, β = 18 mg/dL, p = 3 × 10^-20), and COL18A1 for triglycerides (G111R, 1.9% frequency, β = 16%, p = 4 × 10^-11). With the Exome Chip rare variant genotyping array, we have discovered novel associations between rare variants and lipid levels.

African ancestry is associated with hypertensive cardiomyopathy in Brazilian patients with heart failure. C. Friedman1, M. Cardena1, A. Ribeiro-Santos2, S. Santos2, A. Mansur2, A. Pereira2, 1) Dept of Legal Medicine, Ethics and Occupational Health, Medical School, University of São Paulo, São Paulo, Brazil; 2) Laboratory of Human Genetics and Medicine, Federal University of Pará, Belém, Pará, Brazil; 3) Department of Cardiology, Laboratory of Genetics and Cardiovascular Heart Institute, Medical School, University of São Paulo, São Paulo, Brazil.

Heart failure (HF) is the final pathway for many diseases affecting the heart; it is recognized as the main risk factor for early morbidity and mortality caused by cardiovascular disease. HF is a complex trait resulting from a combination of genetic, environmental, lifestyle and ethnic background, being more prevalent in black individuals. It is known that self-declared ethnicity is not a good method to ethnic classification, especially in admixed populations where there is a little correlation between self-declared skin color and genetic background. Herein, we investigate the association between genomic ancestry and mitochondrial haplogroups (mtDNA-hg) with different heart failure etiologies in 503 Brazilian patients. Mitochondrial haplogroups were obtained by sequencing the control region of mtDNA. Genomic ancestry was obtained using a set of 48 Ancestry Informative Markers (AIMs) inDELDs type. Estimation of global ancestry proportions analyses was done using Structure v.2.2; statistical analysis was done with SPSS 14.0. Different etiologies were observed with our sample: hypertensive (28.6%), ischemic (28.4%), valve (15.1%), chagasiasis (11.9%), idiopathic (9%) and others (6.1%). In the total sample, African mtDNA-hg was the more prevalent (46%), followed by Amerindian (28.2%) and European (25.4%) mtDNA-hg. Genomic ancestry showed the major contribution being European, with 57.5% (±22.1%; 95% CI 51.3-63.7), followed by African (28.3%) and Amerindian (18.2%) (p = 0.065, OR 0.67 [95% CI (1.96-18.9)]). Additionally, we found two new gene associations in the African ancestry haplogroup mtDNA-hg. The analyses of associations with different etiologies showed that individuals belonging to African mtDNA-hg and those with greater contribution of African genomic ancestry present increased risk for hypertensive cardiomyopathy (OR 2.04; 95% CI 1.00-4.12, p = 0.048), CHF (OR 2.28; 95% CI 1.18-4.41, p = 0.012) and ACHF (OR 2.70; 95% CI 1.03-7.18, p = 0.045). Furthermore, we found evidence of association between African ancestry (mtDNA-hg and AIMs) and hypertensive cardiomyopathy in HF patients. Besides, this study support the usage of using AIMs and mtDNA-hg to study the genetic complexity of complex diseases in admixed populations, where skin color is not a good indicator of an individual's genetic ancestry. Supported: FAPESP, LIM40-HC.
2243F
Incidence of aortic root dilatation in patients with 22q11 deletion: The 3-year experience of a VCF specialty clinic. C.B. Hils1, S.L. Dugan2, M.E. Pierpont2, J.D. Sidman3, 1) Children's Heart Clinic, Minneapolis, MN; 2) Genetics Division, Children's Hospitals and Clinics of Minnesota, Minneapolis, MN; 3) Division of Pediatric Genetics and Metabolism, University of Minnesota, Minneapolis, MN.

Background: Congenital cardiac malformations are commonly present in patients with 22q11 deletions. Frequent malformations include cono-truncal defects such as tetralogy of Fallot, interrupted aortic arch, vascular rings/ right aortic arch defects. Aortic root dilatation is also seen in this population, but the incidence and association with specific cardiac diagnoses has not been well-defined. Method and Results: A retrospective chart review was performed for all patients seen at the Velocardiocarbohydrate Syndrome Specialty Clinic at the Children's Hospitals and Clinics of Minnesota from 2009-2012. A total of 78 individual patients were identified (age range newborn to age 29 years). Echocardiogram reports and images were reviewed for each patient. Fourteen patients were referred from outside institutions and aortic root dimensions were not available. All available echocardiograms were reviewed for each patient by a pediatric cardiologist. A total of 179 echocardiograms performed on the remaining 64 patients (range 1-20 studies per patient) were diagnostic for determining the aortic root dimensions at the level of the sinuses of Valsalva. Aortic root dimensions were calculated for each diagnostic study. Thirty-two (50%) of the patients were found to have at least one study in which their aortic root dimensions z-score exceeded a valve of +2. Incidence by cardiac diagnosis varied, with the highest incidence found in patients with tetralogy of Fallot (77%); vascular ring (70%); and aortic arch defects. Aortic root dilatation was observed in aortic arch defects with structurally normal hearts were found to have aortic root enlargement on at least 1 echocardiogram. Several patients were found to have aortic root dilatation dimensions above normal limits with subsequent normalization of the root size for their body surface area; in others, the aortic root enlargement occurred over time. None of the patients in this cohort have required aortic root replacement or have had an aortic root dissection to date. Conclusions: Aortic root enlargement is a common finding in patients with 22q11 deletions, even in the presence of structurally normal hearts. Acute aortic root dissec- tion has been reported only rarely in patients with 22q11 deletions, but additional research is needed to determine the incidence of progressive root enlargement with time and possible need for ongoing screening studies in this population.

2244F
Erythrocyte sedimentation rate (ESR) in ischemic stroke and its relation with stroke risk factors. A. Kaur1, A. Upali2, K. Kaur1, 1) Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 2) Uppal Neuro Hospital, Amritsar, Punjab, India.

Elevated ESR level is considered to be reliable marker of chronic and acute inflammation. Inflammation is hallmark of ischemic stroke. Cardiovas- cular risk factors, chronic and acute inflammation appears to be linked closely to each other and to aggravate the occurrence of various thrombotic events involved in pathophysiology of ischemic stroke. Objective: To investi- gate the relationship between ESR levels and various traditional risk factors in ischemic stroke. Methodology: The study included total of 107 ischemic stroke patients hospitalized in Uppal Neuro hospital, Amritsar (Punjab), diagnosis being confirmed with MRI. Any history of hypertension, diabetes, CAD and atrial fibrillation (AF) was assessed. The patients were classified into three groups according to ESR values measured within 24 hours of admission. Results and Conclusion: A total of 107 ischemic stroke patients including 61 males and 46 females were evaluated for ESR levels, hemoglo- bin (HB) content, TLC, neutrophils and lymphocyte count. Patients with ESR >10mm/hr were included in group I (n=49) with mean age 60 ±12.2. Group II (ESR 11-25) included 72 patients with mean age 65 ±12.2 and patients (n=1) in group III with ESR levels ≤ 28mm/hr had mean age 67 ±12.4. Group I showed higher number of males (79.2%) but Group III with abnormally high values of ESR included most of females (90.9%) vs group II (43.1%) and group I (20.8%). Comparison among three groups with respect to age, HB content, TLC, neutrophils and lymphocytes was carried out with ANOVA (univariate). The difference of mean HB levels among three groups have been found to be statistically significant (F=20.02, p<0.001). No statistical difference was observed among three groups with respect to age (F=2.29, p=0.106), TLC (F=0.78, p=0.46), neutrophils (F=2.36, p=0.09) and lymphocytes (F=2.36, p=0.09). In terms of stroke risk factors including hypertension, diabetes, CAD, no significant difference was observed. It has been shown that patients with low Hb content while on other hand they are not significantly related with various traditional stroke risk factors, may be due to small sample size. Large sample size (being studied) could provide some useful information regarding the relationship between ESR and stroke risk factors.

2244W
Alpha cardiac actin mutation co-segregates with hypertrophic and dilated cardiomyopathy as well as being associated with conduction disease. J. Mcraigh1,2, L. Hunt1, M. Christiansen1, J. Atherton1,2, 1) Gen Hinth Queenslt, Royal Brisbane & Women’s Hosp, Brisbane, Queensland, Australia; 2) University of Queensland, School of Medicine, Brisbane, Queensland, Australia; 3) Statens Serum Institute, Copenhagen, Denmark; 4) Department of Cardiology, Royal Brisbane and Women’s Hospita- l, Brisbane, Queensland, Australia.

Mutations in alpha cardiac actin (ACTC1) have been rarely reported in nonsyndromic familial dilated cardiomyopathy (DCM) and hypertrophic car- diomyopathy (HCM). We report a family with both HCM and DCM associated with conduction disease requiring pacing that co-segregates with an ACTC1 missense mutation. A 40 year old woman with HCM presented for genetic testing. Her father was also said to have HCM and died whilst awaiting cardiac transplantation. She was aware that 2 of her paternal uncles were also said to have cardiomyopathy. ACTC1 missense mutation (c.76 G>A) was identified on genetic screening. Further family studies were suggested to clarify the pathogenicity of the change. One uncle initially presented at the age of 51 years with atrial fibrillation associated with normal LV size, wall thickness and systolic function. He subsequently developed symptomatic atrioventricular (AV) conduction disease, dilated cardiomyopathy and clinical heart failure at 64 years of age and pacing was performed in view of 3rd- degree AV block (despite no negative chronotropic agents). LMNA mutation screening was negative. He also had the ACTC1 missense mutation. Another paternal uncle diagnosed with HCM also had the ACTC1 missense mutation. Conclusion: This case illustrates the phenotypic variation associ- ated with sarcomere mutations. ACTC1 mutations have been previously reported in a small number of HCM and DCM families, however it has not been previously reported in association with conduction disease.

2245F
Thoracic aortic aneurysm (TAA) is a subclinical disease and some cases are associated with aortic dissection and sudden death. TAA can occur as an isolated nonsyndromic condition or in conjunction with inherited connective tissue disorders such as Marfan syndrome. The genetic basis of isolated TAA is identifiable in approximately 20% of cases with up to 14% attributable to mutations in smooth muscle actin (ACTA2). We identified 17 year-old identical twin brothers with severe fusiform TAA with Z-scores at least +7.0 and +7.7, respectively, and ascending aorta Z-scores >+8.0 and +7.7, respectively, that progressed to multiple dissections. Twin 1 presented with dissection of an abdominal aortic aneurysm (AAA). Both brothers had marked auraldysrrhymal dilation of the internal carotid arteries and pulmonary embolism. Eye findings were notable for mydriasis. Molecular testing identified a de novo ACTA2 p.Lys328Asn heterozygous mutation. Testing of FBN1, TGFBR1, TGFBR2, MYH11 was normal. Histologic analysis of thoracic and abdominal aorta samples showed architectural abnormalities including smooth muscle and elastic fiber fragmentation. Smooth muscle actin, the protein encoded by ACTA2, is misexpressed in both thoracic and abdominal aorta tissue samples. Alpha cardiac actin mutation co-segregates with hypertrophic and dilated cardiomyopathy as well as being associated with conduction disease.

2246F
Thrombus formation in patients with TAA is characterized by smooth muscle and TGFβ signaling abnormalities in conjunction with a potential role for inflammation in the thoracic aorta.
2247W
1) Internal Medicine, University of Texas Health Science Center at Houston, Houston, TX; 2) Pediatrics, University of Texas Health Science Center at Houston, Houston, TX; 3) St. George’s University of London, London, UK; 4) National Obstetrical and Cardiac Vascular Center Research Institute, Suita, Japan; 5) Stanford University, Stanford, CA.

ACTA2 mutations are the most common cause of familial thoracic aortic aneurysms and dissections (TAAD) responsible for 10-14% of the disease and predispose to occlusive vascular diseases leading to stroke and coronary artery disease (CAD) with strong association of specific mutations with vascular diseases. Disease risk estimates are important for genetic counseling and surveillance of at-risk patients, but are not available for ACTA2 mutations. Using data collected from the largest cohort of individuals with a spectrum of ACTA2 mutations (n = 217), we determined Kaplan-Meier estimates of failure function for TAAD (aortic aneurysms requiring surgical repair or aortic dissections), CAD or myocardial infarct (MI), and stroke for all mutations and specific mutations occurring in at least 2 unrelated families. By age 70 years, 85% (95% CI = 74-93%) of individuals had TAAD. Percentage of individuals who had TAAD from birth to oldest age of follow-up was 100% by age 25 years for Arg179, 100% by age 45 years for Arg198, 93% (95% CI = 74-100%) by age 63 years for Arg258, 85% (95% CI = 53-99%) by age 62 years for Arg118, 75% (95% CI = 59-86%) by age 70 years for Arg149, 70% (95% CI = 37-96%) by age 76 years for Arg212, 68% (95% CI = 44-90%) by age 65 years for Arg39, 67% (95% CI = 32-95%) by age 56 years for Gly160, and 60% (95% CI = 17-99%) by age 77 years for Arg185. Using the logrank test, risk for TAAD was significantly higher for mutations altering Arg179 (p<0.001) and Arg258 (p = 0.017) compared to other ACTA2 mutations. Eighteen percent (95% CI = 12-28%) of individuals had early onset CAD or MI and 13% (95% CI = 9-20%) had early onset stroke. Mutation-specific risk estimates were also calculated for early onset stroke and CAD or MI. Risk for early onset stroke was significantly higher for mutations altering Arg179 (42%, 95% CI = 17-80%, p<0.001) and Arg258 (41%, 95% CI = 20-71%, p<0.001) but lower for Arg149 and Arg39 (10%, 95% CI = 0-38% and 16%, 95% CI = 0-35%, respectively). These findings demonstrate variable disease risk specifically for ACTA2 mutations. Further studies are needed to provide detailed risk estimates that will aid in genetic counseling, vascular imaging and management of patients, and provide a guide for assessing aortic disease risk for other genes that predispose to this disease.

2248T

Cathepsin A is lysosomal enzyme that makes complex with α-neuraminidase and β-galactosidase. In the complex, it activates α-neuraminidase and stabilizes β-galactosidase. Cathepsin A is an acidic serine carboxypeptidase with deamidase and esterase activities at neutral pH. It catalyzes the hydrolysis of C-terminal peptide bond which potentially affects vasoactive peptide actions in normal and pathological conditions. In order to elucidate the biological function of Cathepsin A in vivo, we previously generated knock-in mouse model with catalytically inactive Cathepsin A protein. We showed that Cathepsin A-/- mice have significantly increased arterial blood pressure and higher level of endothelin-1 secreted from cultured brain cells. In this work, we present our immunochemical study of vasoactive peptides in brain and visceral organs from Cathepsin A-/- mice and their wild type siblings. We showed higher level of endothelin-1 and substance P accumulation in brain from 3 and 6 months old Cathepsin A-/- mice compared to wild-type mice. However, we found that Cathepsin A deficiency did not result in significant accumulation in kidney, liver and lung. The detailed analysis of other vasoactive peptides including angiotension I, bradykinin and oxytocin in brain and visceral organs of Cathepsin A-/- mice will be valuable to understand better the physiological role of this lysosomal enzyme in the modulation of vasoactive peptides and complex processes such as regulation of blood pressure.

2249F
The Yield of Clinical Genetic Testing for Isolated Left Ventricular Non-compaction (LVNC). A.R. Shikany, A.S. Parrott, S. Ware, E.M. Miller. The Heart Institute, Cincinnati Children’s Hospital, Cincinnati, OH.

Left ventricular noncompaction (LVNC) is a distinct form of cardiomyopathy characterized by two layers of compacted and noncompacted myocardium with prominent trabeculations in the noncompacted layer. LVNC is clinically heterogeneous with subtypes including isolated LVNC (LVNC with otherwise normal left ventricular size, thickness, and function) and non-isolated LVNC (LVNC co-occurring with another form of cardiomyopathy). Mutations in 11 genes known to cause cardiomyopathy have been identified in 40% of cases of LVNC, not defined by subtype. This study sought to investigate the yield of genetic testing among patients with isolated LVNC. This retrospective chart review included patients evaluated in the multidisciplinary Cardiomyopathy Clinic at Cincinnati Children’s Hospital between July 2009 and December 2012. Clinical data was collected on 169 patients from 144 families for whom LVNC was confirmed by a cardiologist. Seventy-one probands had a diagnosis of isolated LVNC and 73 probands had non-isolated LVNC. Of individuals with isolated LVNC, 46% (33/71) had genetic testing, of which 12% (4/33) had a disease-causing mutation. The yield of genetic testing in the non-isolated LVNC group was 35% (16/48). A variant of unknown significance (VUS) was identified in 21% (7/33) and 17% (8/48) of the isolated LVNC and non-isolated LVNC subtypes respectively. Of the individuals with isolated LVNC who had a disease-causing mutation, all (4/4) had a reported family history of LVNC, cardiomyopathy, or sudden cardiac death. Given the low yield of testing and risk for a VUS, results from this study suggest that testing for isolated LVNC may be of limited clinical utility in the context of a negative family history. Screening of first-degree relatives should be considered prior to genetic testing to confirm family history status. Healthcare providers should consider LVNC subtype in their provision of counseling regarding the benefits and limitations of genetic testing for LVNC.

2250W
Causative analysis of markers related to PR interval properties and atrial fibrillation after cardiac surgery. M.I. Sigurdsson1, J. Muehlschlegel1, A. Fox1, M. Heydarpour2, P. Lichtner1, T. Meltinger2, C. Collard2, S. Sherman2, S.C. Body1. 1) Anesthesia, Perioperative and Pain Medicine, Brigham and Women’s Hospital, Boston, MA; 2) Institut für Klinische Genetik, University of Heidelberg, Germany.

Several genetic loci linked to atrial fibrillation (AF) have also been associated with PR interval timing. In a subset of patients with AF, we hypothesized that variants associated with PR interval timing could impact AF. We genotyped 28 SNPs associated with PR interval properties and AF in 1314 patients undergoing coronary artery bypass surgery. In this analysis, 21 SNPs with P-values < 0.05 that were associated with both PR interval properties and AF were considered for further analysis. We tested for association with properties of PR interval. Genetic variants associated with AF or PR interval implicating a mechanistic role. We determined the association between genetic variants and AF following coronary artery bypass surgery by pathways including, and independent, of the PR interval utilizing methodology that adjusts the association between a genetic marker and AF for properties of PR interval. Genetic variants associated with AF or PR interval from literature were tested for association with AF and PR interval in a cohort of 1314 patients undergoing coronary artery bypass grafting. Those associated with both were analyzed with association with poAF independent of the PR interval. Patients with AF had longer pre-operative PR interval, maximum PR interval and maximum change in PR interval and change in PR interval from pre-operative value. Two regions on 1q21 and 4q25 had several markers associated with both AF and the PR. All of those associated with AF independent of the PR interval. In contrast rs740178 on chromosome 16 was associated with AF only via the PR interval. Our results highlight the utility of studying the molecular background of cardiac pathophysiology in a thoroughly phenotyped cohort of patient undergoing cardiac surgery.
2251T
Clinical and mutational spectrum in patients with cardiac glycogenosis related to mutations in the prkag2 gene. J. Thevenon1, G. Laurent2, P. Charon3, P. Laforêt4, A. Millaire4, D. Klug5, L. Gouya6, C.A. Maurgeon1, S. Kacet7, J.C. Eicher7, X. Jeunemaitre6, M. Desnos6, E. Bieth7, P. Bouvagnet5, D. Duboc11, L. Martin12, P. Réant13, F. Picard13, C. Bonithon-Kopp14, E. Gautier10, C. Binquet14, C. Thauvin-Robinet15, J.E. Wolf1, L. Faivre1, P. Richard15, 1) Centre de Génétique et de Centre Référence Anomalies du Développement et Syndromes Malformatifs, Hôpital d’Enfants, CHU Dijon et Université de Bourgogne, France; 2) Centre de Cardiologie Clinique et Interventionnelle, Hôpital du Bocage, Centre Hospitalo-Universitaire de Dijon, Dijon, France; 3) AP-HP, Département de Génétique, Centre de référence des maladies cardiaques héréditaires, INSERM UMR S956, UPMC Université Paris 6 ; Hôpital de la Pitié-Salpêtrière, Paris, France; 4) Centre de Référence de pathologie neurovasculaire Paris-Est, Groupe Hospitalier Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Paris, France; 5) Hôpital Cardiologique, CHRU de Lille, France; 6) INSERM U773, Centre de Recherche Biomédicale Bichat Beaujon CRBB, Université Paris 7 Denis Diderot, site Bichat, Paris, France; 7) Université Lille Nord de France, USD, EA 1036, F-59000 Lille, France; 8) Department of Pathology, Lille University Hospital, F-59000 Lille, France; 9) Hôpital Purpan, Département of Medical Genetics, Toulouse, France; 10) Laboratoire Cardiogénétique, Groupe Hospitalier Est, Hospices Civils de Lyon, Lyon, France; 11) Service de Cardiologie, Hôpital Cochin, Paris, France; 12) Laboratoire de Génétique, CHU, Lyon, France; 13) Department of Surgical Cardiology, Haut Leveque Hospital, Pessac, France; 14) Centre d’investigation clinique -épidémiologie clinique/ essais cliniques, CHU, Dijon, France; 15) APHP, UFI Cardiogénétique et Myocardiologie. Service de Biochimie Métabolique, Groupe Hospitalier Pitié- Salpêtrière 47-83 boulevard de l’Hôpital, 75651 Paris cedex 13, France.

Background: PRKAG2 mutations are responsible for an autosomal dominant muscular glycogenosis associating hypertrophic cardiomyopathies (HCM), dilatative cardiomyopathies (DCM), and progressive atrio-ventricular block (AVB) and sometimes sudden cardiac deaths (SCD). Objectives: To describe the time dependent clinical occurrence of clinical manifestations of the disease. To define the indications for PRKAG2 mutation screening.

Methods: A cohort of 34 patients from 21 families: 1) 17 patients with AVB and 23 patients with HCM from a national French cohort of 140 patients tested for PRKAG2 gene since 2001. The pick-up rate varied from 0% when the indication was an isolated HCM to 20% when associated with a VPE isolated short PR interval or WPW syndromes. Aside from the recurrent p.Arg302Gln (48%), we identified four new mutations. Time dependant risk of complication occurrence was estimated using a Kaplan-Meier method. At 40 years of age the chances to present with a VPE was 70% (95%-CI: 50%-87%), with a HCM was 61% (95%-CI: 41%-81%), and to die from a SCD was 20% (95%-CI: 8%-42%). At the 66 years, 32% of the patients required an implantable cardiac defibrillator and 2 required heart transplant at 55 and 60 years. No significant genotype-phenotype correlation has been found. However p.Arg302Gln mutation was less of associated with a HCM and no WPW syndrome was diagnosed in the p.Arg302Gln group. WPW radiofrequency ablation was associated with a complete AVB in 4/6 patients. Conclusion: Because of the risk of SCD, PRKAG2 screening may be useful in case of VPE with or without HCM.

2252F
Genetic profiling supports the causal role of type 2 diabetes, and fasting insulin and glucose in cardiovascular diseases. E. Tikkanen1, A.S. Havulinna1, V. Salomaa2, S. Ripatti2,3,4, 1) Institute for Molecular Medicine Finland FIMM, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Hjelt Institute, University of Helsinki, Helsinki, Finland; 4) Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Epidemiological studies have identified several traits associated with cardiovascular disease (CVD), but few of these have been shown to be causal risk factors. Genetic information obtained from genome-wide association studies (GWAS) could be used to assess the causality of risk factors on disease. The aim of this study was to examine the causal biological pathways for CVD by genetic profiling of several risk factor traits. By utilizing the information from published GWAS, we generated genetic risk scores (GRSs) for blood lipids, obesity, blood pressure, inflammation and type 2 diabetes-related traits in Finnish case-cohort dataset (N=1974) with extensive clinical and follow-up data. The GRSs were calculated by summing the number of risk alleles and dividing the sum with the number of SNPs, after which each GRS were standardized (mean=0, SD=1). Associations between the GRSs and cardiovascular events were tested with Cox proportional hazards models and two-sided P<0.05 were considered to be statistically significant. To evaluate whether the GRS effects are governed by the proportion of heritability (h2) explained by the genetic variants, we estimated h2 from genome-wide data of Finnish cohorts (N(max)=14,073) and calculated the proportion of h2 explained by the GRS for each trait. We found that the GRSs for insulin, type 2 diabetes, triglycerides and total cholesterol increased the risk for incident coronary heart disease in an analysis adjusted for age and gender. The GRSs for type 2 diabetes and glucose was associated with incident stroke. In a multivariable analysis accounting for all risk factor traits and genetic risk scores, the GRS for insulin remained associated with coronary heart disease (HR=1.44, 95% CI 1.15-1.79). The GRSs for insulin and glucose account for 19% and 45% of the trait heritability, respectively. The proportion the GRS explains of the h2 was not correlated with the association of trait-based GRS with CVD risk (r<0.005). We conclude that these results support the causal role of type 2 diabetes-related traits for cardiovascular disease.

2253W
Left ventricular non-compaction, Ebstein’s anomaly and autosomal dominant polycystic kidney disease: A novel association and review of the vascular features of ADPKD. L. Zahavih1, A. Dipchand1, S. Bowdin1, 2, 1) The Hospital for Sick Children, Toronto, Canada; 2) University of Toronto, Toronto, Canada.

Autosomal dominant polycystic kidney disease (ADPKD) is the most common inherited renal disease and one of the most common genetic disorders with a prevalence of 1/400-1/1000. There are two known genes associated with ADKD, PKD1 and PKD2 accounting for about 90% of cases. Left ventricular non-compaction (LVNC) is characterized by persistent fetal myocardial disarray leading to trabeculations and a thickened 2-layer myocardium. In some patients, this can result in heart failure. Familial recurrence of LVNC has been reported in about 20% of cases. Genetic causes include sarcomere gene mutations, Barth syndrome and chromosomal variants. We present a case of a 2 year old boy who was diagnosed with Ebstein’s anomaly and echogenic kidneys prenatally. He presented with congestive heart failure at 3 weeks of age and was then diagnosed with LVNC. At 9 months he was found to have bilateral renal cystic disease. Family history is significant for PKD, bicuspid aortic valve and aortic aneurysm in his maternal grandmother as well as PKD and valvular disease in extended maternal relatives. Testing of the MYH7 gene was organized given the association between LVNC, Ebstein’s anomaly and MYH7 mutations and was normal. He was subsequently found to have a well described mutation in PKD1 (p.Q4042X). Vascular abnormalities are frequent in patients with ADPKD, particularly hypertension and valvular disease. Common valvular anomalies include mitral valve prolapse, mitral or tricuspid valve regurgitation and aortic root dilatation. There have been reports of left ventricular hypertrophy in association with ADPKD, including in those who are normotensive. Previous studies have also noted an increased prevalence of dilated cardiomyopathy in patients with ADPKD. To date, there have been three reported cases of LVNC in patients with ADPKD, two adults and one pediatric case. The pathogenesis of LVNC in ADPKD is currently unknown, however a two-hit model of cyst formation has been hypothesized. This model proposes that PKD null cells induce apoptosis of surrounding normal cells leading to cyst formation. Cardiac defects and abnormal myocardial trabeculations have been described in PKD1 mutant mice. A potential mechanism for LVNC in patients with ADPKD could involve induced apoptosis of the myocardium leading to persistent trabeculations. Further studies of the cardiac phenotype in ADPKD are required to assess the relationship between ADPKD and LVNC.
2254T

Background: Copy number variations (CNVs) were initially recognized as genomic non-functional rare rearrangements but currently around 180,000 CNVs are identified along the human genome and contribute approximately to 5% of genome variation among individuals. CNVs can impact phenotypes by gene dosage modifications. Besides, Tetralogy of Fallot (ToF) is present in 1/3000 live births and accounts for 60% of conotruncal malformations. Associated mutations have been described in early developmental genes including NKX2-5, NOTCH1, TBX1, JAG1, NOTCH2. Studying nuclear families, Greenway et al. identified seven CNVs associated with an increased risk for ToF, three of which involved genes previously associated with heart disease: TBX1, NOTCH1 and JAG1. On the other hand, MLPA is a high performance and low cost technique that allows the analysis of over 40 target genome sequences with a single pair of primers in a single reaction. Aim: To identify the frequency and type of certain CNVs related to congenital heart disease, including those present in specific genes (GATA4, NKX2-5, TBX5, BMP4, CRELD1) and other 48 genomic sequences distributed throughout the chromosomal regions: 10p14, 4q35, 17p13, 22q11 and 22q13 by means of MLPA technique in a sample of Mexican patients with ToF, their parents and a control group. Material and Methods: 28 family trios in which the propositus had the diagnosis of ToF and 28 healthy controls were analyzed with the MLPA technique using the P311 and P250 kits from MRC-Holland®. Real Time-PCR and FISH with specific probes were used to validate the results. Results: One propositus showed a de novo CNV in the TBX1 gene, involving exons 2 and 7, both parents were normal. We did not observe any CNV in the control group. Conclusions: We identified a TBX1 gene deletion in 1/28 cases with ToF, which was not detected by conventional karyotype or FISH analysis for the 22q11 region. The proportion of cases with de novo CNVs in our sample is according to the observed by other research group. The TBX1 variant in our patient could explain the presence of malformations, particularly in early cardiac development. The MLPA technique allows the detection of chromosomal alterations such as microdeletion and microduplication not identified by conventional karyotype or FISH analysis, at a low cost and with a fast performance.

2255F

The leading cause of death in the world is coronary artery disease (CAD). There have been numerous risk factors identified for the development of CAD and it’s main complication myocardial infarction (MI), including high plasma lipid concentrations, high blood pressure, smoking, diabetes, and markers of inflammation. We hypothesized that epigenetic markers may add to the growing etiology of myocardial infarction and provide additional biomarkers that can be used to predict MI in individuals with CAD. METH-ODS: The DNA was extracted from whole blood and analyzed for methylation changes using the IllumaHumanMethylation450 Beadchip (485,577 CpG sites). Statistical analysis was completed on 395,899 CpG sites (Pardek Genomics Suite) comparing males with stable CAD (n = 7) and CAD and MI (n = 7). Any CpG sites with an average Beta value less than 0.1 was removed due to the sensitivity of the Beadchip assay. RESULTS: After the scrubbing of data, 7,131 CpG sites were significantly different compared to stable CAD and males with CAD with MI (p < 0.05). There were 1,710 CpG sites that showed a higher beta value in CAD group compared to CAD with MI. The top ranked CpG site by p-value (p = 2.38 × 10-06; cg068583846) was located in the promoter of the gene SERPINE1 and showed 81 percent higher beta value in individuals with stable CAD. CONCLUSIONS: This work provides new insight into the biological differences between individuals with stable CAD and CAD with a myocardial infarction. Additionally, we can utilize epigenetic markers to discern novel pathways involved in MI disease etiology.

2256W
Evaluation of the Familiality of Cardiovascular Diseases among Patients in a Large Healthcare System. S. Knight1, B. Home2, 1) Intermountain Heart Institute, Intermountain Medical Center, Murray, UT; 2) Division of Genetic Epidemiology, University of Utah, Salt Lake City, UT.

Objective: To determine the familiality of cardiovascular diseases (CVD) Methods: The Intermountain Genealogy Registry, which contains genealogies for 700,000 patients (pts) in the Intermountain Healthcare System, was used. One-way and two-way genealogy index of familiality (GIF) were generated. Results: The overall pfs GIF (pGIF) was 0.443. The GIF for patent foramen ovale (PFO) was >40% and myocardial infarction (MI) was >10% larger than the pGIF. No two-way GIF were >10% larger than the pGIF. Conclusions: PFO and MI had a stronger familial component than other CVD. Narrowing phenotypes may increase the GIF.

<table>
<thead>
<tr>
<th>GIF</th>
<th>MI</th>
<th>HF</th>
<th>CAD</th>
<th>HTN</th>
<th>PFO</th>
<th>Valve</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>0.488</td>
<td>0.462</td>
<td>0.470</td>
<td>0.453</td>
<td>0.474</td>
<td>0.466</td>
</tr>
<tr>
<td>HF</td>
<td>0.471</td>
<td>0.457</td>
<td>0.446</td>
<td>0.453</td>
<td>0.458</td>
<td>0.460</td>
</tr>
<tr>
<td>CAD</td>
<td>0.476</td>
<td>0.450</td>
<td>0.465</td>
<td>0.460</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HTN</td>
<td>0.461</td>
<td>0.451</td>
<td>0.447</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PFO</td>
<td>0.628</td>
<td>0.459</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valve</td>
<td>0.476</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2257T
Genome-wide linkage and positional association study of blood lipid phenotypes: The GenSalt study. C. Li1, J.E. Hixson2, L.C. Shinnim3, D.C. Rao4, D. Gu5, J. He6, T.N. Kelly1. 1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA; 2) Department of Epidemiology, University of Texas School of Public Health, Houston, TX, USA; 3) Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA; 4) State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

The genetic mechanisms underlying lipid levels remain largely unknown. We conducted a genome-wide linkage scan and positional association analysis of blood lipid phenotypes among Han Chinese participants of the Genetic Epidemiology Network of Salt Sensitivity (GenSalt) study. Overnight fasting blood samples were drawn by venipuncture at a baseline examination among 1,881 GenSalt participants from 633 families. Total cholesterol (TC), high density lipoprotein cholesterol (HDL-C) and triglycerides (TG) were analyzed enzymatically using commercially available reagents. Low density lipoprotein cholesterol (LDL-C) levels were calculated using the Friedewald equation. Lymphocyte DNA samples were used for genotyping microsatellite markers and single nucleotide polymorphisms (SNPs). Multipoint quantitative trait linkage-analysis was performed for each of the TC, HDL-C, LDL-C, and log-transformed TG phenotypes using SOLAR software. Additive effects of alleles between each SNP in identified linkage regions (Munh-2) and the lipid phenotypes were assessed using a mixed linear regression model to account for family dependencies. The false discovery rate (FDR) method was used to adjust for multiple testing. Phenotypes were adjusted for age, gender, body mass index and field center in all analyses. Suggestive linkage signals were identified at 4p15.1-4p14 [maximum multipoint LOD score (MML)=2.45 at 4p15.1] for the LDL-C phenotype. Follow-up association analyses in this region revealed a significant association between novel marker rs2995976 and LDL-C after adjustment for multiple testing (P=3.73×10-10; FDR-Q=0.03). Median LDL-C levels for rs2995976 genotypes G/G, A/G, and A/A were 88.36 (82.56 to 94.17), 93.06 (90.62 to 95.50), and 97.69 (95.97 to 99.41) mg/dL, respectively. SNP rs2995976 lies upstream PGM2, a gene indirectly involved in pathways of fatty acid production and energy generation. Furthermore, we observed suggestive linkage at 2p11.2-2q12.3 (MML=2.22 at 2q11.2) and 11q25 (MML=2.21 at 11q25) for the log-transformed TG phenotype. Follow-up analyses did not reveal any significant association signals in this region. In summary, genomic regions on chromosomes 2, 4 and 11 may harbor important susceptibility loci for lipid phenotypes. In addition, a novel variant upstream of the PGM2 gene was significantly associated with LDL-C. Further study is needed to confirm these findings.
2258F
Genome-wide linkage and regional association study of blood pressure response to cold pressor test in Chinese: the GenSalt study. X. Yang1, T.N. Kelly2, X. Wu1, J.E. Hixon3, J. Chen1, J. Cao1, J. Li1, L.C. Shimmim14, J. Huang1, D.C. Rao1, J. He2, D. Gu1. 1 State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; 2 Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA; 3 Department of Epidemiology, University of Texas School of Public Health, Houston, TX, USA; 4 Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA.

The cold pressor test (CPT) has been documented to evaluate cardiovascular reactivity to stress. However, genetic mechanisms underlying blood pressure (BP) response to CPT remain unclear. We performed a genome-wide linkage scan and regional association analysis to identify genetic determinants of BP response to CPT. A total of 1,998 Han Chinese participants from the Genetic Epidemiology Network of Salt Sensitivity (GenSalt) study completed the CPT. BP measurements were obtained before and after the participants immersed their hands in iced-water for 1 minute. BP response to the CPT was defined as BP at time zero after iced-water immersion minus the pre-test BP. Multipoint quantitative trait linkage analysis was performed using SOLAR software. The additive associations between single SNPs in linkage regions (LOD>2) were assessed using a mixed linear regression model to account for familial correlations. In addition, gene-based associations under linkage peaks were also examined using the truncated product method. Age, gender, field center, and body mass index were adjusted in all analyses. The results showed that suggestive linkage signals were identified for BP responses to CPT at 20p13-20p12.3, with maximum multipoint LOD scores of 2.38 for systolic BP (SBP) response, 0.13 for diastolic BP (DBP) response and 1.30 for mean arterial pressure (MAP) response. Marker rs3236373, located in the 5' flanking region of the SMOX gene at 20p13, was significantly associated with DBP and MAP responses to CPT (P=4.49×10−5 and 6.52×10−6, respectively). DBP responses (95% CI) for genotypes G/G, G/A, and A/A were 6.84 (6.44 to 7.23), 4.11 (2.83 to 5.40), and 4.49×10−5, respectively. DBP was significantly associated with DBP and MAP responses to CPT (P=2.09E-4 - 2.69E-2), and cell death and survival (P=3.17E-4 - 4.53E-2).

2259W
Whole-genome expression profile of calcified bicuspid and tricuspid aortic valves. S. Guauque-Olarte1, N. Gaudreault1, P. Pibarot1, P. Mathieu1, Y. Bossé1,2, Centre de recherche Institut universitaire de cardiologie et de pneumologie de Québec, Quebec, Canada; 2) Department of Molecular Medicine, Laval University, Quebec, Canada.

BACKGROUND: Calcific aortic valve stenosis (AS) is a fatal disease with no medical treatment other than surgical intervention. Bicuspid aortic valve (BAV) is present in 1-2% of the population. Patients with BAV have increased risk of AS and develop symptoms 10-15 years younger than patients with a tricuspid valve (TAV). The objective of this study was to identify genes differentially expressed between BAV and TAV with or without calcification.

METHODS: Twelve calcified BAV and 12 calcified TAV were explanted from white male patients who underwent aortic valve replacement surgery. All valves had the same degree of fibro-calcalcic remodeling. Non-senitonic TAV (n=6) were taken from 12 white males who underwent heart transplantation. The gene expression profile of each valve was measured with the Illumina HumanHT-12 v4 Expression BeadChip. Normalization and quality controls were performed with the lumi package in R. Gene expression levels were compared between the three groups of valves (BAV, TAV, and controls) using the Significance Analysis of Microarrays program. A false discovery rate of 5% and a fold change cut-off of 2 were used. Pathway analysis was performed using Ingenuity Pathway Analysis. The microarray results will be compared with RNA-sequencing data from the same valves performed on the Illumina HiSeq2000.

RESULTS: Two up-regulated and 2 down-regulated genes were identified in BAV compared to calcified TAV. For the comparison BAV vs controls, 128 genes were differentially expressed including 80 genes up-regulated in BAV. Compared to controls, 42 genes were up-regulated in calcified BAV and 18 genes were down-regulated. The fold change for these expressed between BAV and calcified TAV were significantly linked to pathways and functions related to AS development and progression including connective tissue disorders (p=2.09E-4 - 6.07E-4), inflammatory disease (p=3.29E-4 - 2.69E-2), and cell death and survival (p=3.17E-4 - 4.53E-2).

CONCLUSIONS: The gene expression profiles of calcified BAV and TAV are highly similar. In contrast, aortic stenosis induced substantial changes in aortic valve gene expression. The results of this study increased our understanding of the molecular mechanisms of BAV and the pathogenesis of calcific stenosis.

Expected outcomes are new therapeutic targets to prevent, slow the development or treat AS in patients with BAV and calcified TAV.

2260T
Peripheral blood microRNA profiles are associated with cardiometabolic disease. L.C. Kwee1, W.E. Kraus1, U. Hidefumi1, H. Toyoshiba2, T. Andou3, E.R. Hauser1, S.G. Gregory1, J. Bain1, M. Muehlbauer1, R. Urrutxarri1, C.B. Newgard1, S.H. Shah1, 1 Duke Institute of Molecular Physiology, Duke University Medical Center, Durham, NC; 2) Pharmaceutical Research Division, TaXeD PharmaCompany, Japan.

MicroRNAs (miRNAs) are short, noncoding RNA molecules that affect gene expression and have been implicated in cardiometabolic diseases. Small-molecule metabolites also are associated with these phenotypes; however, miRNA and metabolite profiles have not been integrated in cardiometabolic disease studies. We examined the relationship between circulating miRNA profiles with metabolites and cardiometabolic phenotypes to identify miRNAs that may mediate cardiometabolic disease.

Fasting plasma samples were obtained from 712 subjects selected from the CATHGEN biorepository of patients referred for cardiac catheterization at Duke University. miRNA levels were assessed using TaqMan RT-qPCR arrays, and the most variable quantile of mean-normalized miRNAs was used for analysis. Lipids, glucose, ketones, free fatty acids, 15 amino acids and 45 acylcarnitines (10 miRNAs, lowest p=3.1×10−6), a factor composed primarily of free fatty acids, (41 miRNAs, lowest p=10−4), short-chain dicarboxylacylcarnitines (53 miRNAs, lowest p=10−2) and with triglyceride levels (4 miRNAs, lowest p=10−3); 36 miRNAs were associated with both phenotypes. The predicated gene target for the most significant differential miRNA profile was YOD1, which is involved in the biosynthesis of unsaturated fatty acids (KEGG). Additionally, multiple miRNAs were associated with long-chain acylcarnitines (10 miRNAs, lowest p=3.1×10−4), short-chain dicarboxylacylcarnitines (36 miRNAs, lowest p=1.7×10−4), a factor composed primarily of asparagine/aspartic acid and glutamine/glutamic acid (5 miRNAs, lowest p=4.2×10−4), and glucose (4 miRNAs, lowest p=3.2×10−4). In one of the largest studies of miRNAs in human cardiometabolic disease, we have identified biologically plausible and novel phenotypes associated with peripheral blood miRNA levels. These miRNAs and their predicted targets are excellent candidates for further analysis as potential regulators of cardiometabolic phenotypes.
2261T

**DYNAMICS OF HSPC SUBTYPES IN NON-HUMAN PRIMATES REVEALED BY A DECADE-LONG CLONAL TRACKING STUDY.**


1) Microbiology, Immunology, and Molecular Genetics; 2) Biostatistics; 3) Pathology and Laboratory Medicine; 4) School of Nursing, University of California Los Angeles, Los Angeles, CA 90095; 5) National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892; 6) Korean Bioinformatics Center, Daejeon, South Korea.

Hematopoietic stem cell and progenitor cell (HSPC)-based genetic therapy to treat HIV/AIDS and other previously incurable diseases is becoming increasingly attractive, however, the potential therapeutic benefits of these cells depend on the function and fate of the HSPCs. We analyzed the long-term behavior of HSPCs in the gut microbiota of macaques and rhesus monkeys over 10 years and found that the HSPC subtypes and their repopulating potential change over time. This information can be used to improve individual treatment options for PFIC patients.

**Results:**

1. **Characterization of HSPC subtypes:** We identified two major HSPC subtypes, type A and type B, based on their surface markers and their ability to repopulate the blood and spleen. Type A HSPCs were more prevalent in young animals and showed higher repopulating potential, while type B HSPCs were more prevalent in older animals and showed lower repopulating potential.

2. **Replication of HSPCs in the gut microbiota:** We observed that the HSPC subtypes changed over time, with type A HSPCs becoming more prevalent in older animals and type B HSPCs becoming more prevalent in younger animals.

3. **Impact of HSPC subtypes on disease:** We found that the HSPC subtypes were associated with different disease outcomes, with type A HSPCs being associated with better outcomes and type B HSPCs being associated with worse outcomes.

4. **Dynamics of HSPC subtypes:** We observed that the HSPC subtypes changed over time, with type A HSPCs becoming more prevalent in older animals and type B HSPCs becoming more prevalent in younger animals.

**Conclusion:**

Our findings suggest that the HSPC subtypes may be useful for predicting disease outcomes and for selecting the best treatment options for PFIC patients.
2264F
Creation of a new mouse model for the mRNA splicing disease Familial Dysautonomia. A. Morini1, P. Dietrich2, I. Dragassis2, M. Salani1, F. Urba3, S.A. Staugast.Philadelphia1, 1) Center for Human Genetic Research, Massachusetts General Hospital/Harvard Medical School, Boston, MA; 2) Department of Physiology, The University of Tennessee, Health Science Center, Memphis, TN.
Recent studies emphasize the importance of mRNA splicing mutations in genetic disease, with some estimates suggesting that around 30% of point mutations disrupt mRNA splicing. Familial dysautonomia (FD) is an excellent disease model for studying new strategies to correct splicing defects. FD is a recessive neurodegenerative disease caused by a splice mutation in the IKBKAP gene which leads to variable skipping of exon 20. We found that the small molecule kinase can correct the IKBKAP splicing defect and increase the amount of normal mRNA and protein in FD cell lines. We have also shown that kinase can increase the level of functional IKBKAP protein in mice following oral dosing in all tissues tested, including brain. Despite these remarkable advances, including our demonstration of in vivo efficacy to increase normal IKBKAP mRNA in the blood of FD patients, we lacked animal models in which to test the effect of targeting mRNA splicing to increase IKBKAP protein on FD phenotype. In order to create a phenotypic model of FD in which we could also manipulate mRNA splicing we introduced an FD transgene (TgFD9), which contains the human IKBKAP gene with the major FD splicing mutation, into the ikbkapΔ/Δ mouse model by sequential mating. The introduction of the human IKBKAP transgene attenuates the severe FD phenotype that we observed in the ikbkap Δ/Δ mouse and recreates the same tissue-specific mis-splicing defect. TgFD9/ikbkapΔ/Δ mice show a reduced growth rate but, unlike ikbkap Δ/Δ mice which die perinatally, most F9D/ikbkap Δ/Δ mice survive post-natally. Methylen blue staining of tongues shows that F9D/ikbkap Δ/Δ mice have a reduction of fungiform papillae. Since several FD-like features are present at birth, we analyzed body weight, superior sympathetic ganglia (SSG), stellate ganglia (SG) and cervical DRGs in E18.5 F9D/ikbkap Δ/Δ and wild type embryos. Already at this stage the F9D/ikbkap Δ/Δ embryos accurately model many features of the human disease. Furthermore, we characterized the mechanism underlying the low levels of endogenous mouse ikbkap expression in ikbkapΔ/Δ mouse embryos. These studies have implications for understanding the pathogenesis of phenotypes in FD patients. The creation of this new model has allowed us to initiate a detailed clinical trial of kinase and will permit testing of other strategies for targeting mRNA splicing.

2265T
Phosphatidylserine: A potential gene modifying therapy for Familial Dysautonomia. M. Salani1, L. Nociforo-Kaufmann2, J. Martinez2, E. Morini1, F. Axelrod3, S. Staugast1. 1) Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA; 2) Dysautonomia Center, New York University School of Medicine, New York, NY.
Familial dysautonomia (FD) is caused by a splicing error in the IKBKAP gene that encodes human Elongator protein 1. In these patients, exon 20 is frequently skipped during mRNA splicing, but cells retain the ability to produce a low level of normal (wild-type) IKBKAP mRNA and normal IKBKAP protein. Phosphatidylserine (PS, Sharp-thought®), an acidic phospholipid, is frequently skipped during mRNA splicing, but cells retain the ability to produce a low level of normal (wild-type) IKBKAP mRNA and normal IKAP protein. Phosphatidylserine has been shown to raise IKAP levels in fibroblast cell lines derived from FD patients. We have shown that PS might safely raise normal IKBKAP mRNA levels in blood from FD patients, opening an exciting potential therapeutic path for the treatment of FD. A more extensive analysis using a larger number of patients and an increased dose of PS is underway. 1 Keren et al., PLOS One. 2010 Dec 29;5 (12): e15884.

2266F
Inhibition of retinoic acid signaling rescues inner-ear defects in a mouse model of CHARGE syndrome. J.M. Skidmore1, E.A. Hurds1, A. Saikakhova1, D.L. Swiderski1, E.D. Sperry2, P.S. Scacheri1, Y. Raphael2, D.M. Martin1,5,6. 1) Department of Pediatrics, University of Michigan, Ann Arbor, MI; 2) Centre for Neuroregeneration, University of Edinburgh, Edinburgh, Scotland; 3) Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH; 4) Department of Otolaryngology, University of Michigan, Ann Arbor, MI; 5) Medical Scientist Training Program, University of Michigan, Ann Arbor, MI; 6) Department of of Human Genetics, University of Michigan, Ann Arbor, MI.
CHARGE syndrome, a multiple congenital anomaly disorder, and in utero vitamin A deficiency or toxicity share many common developmental features including craniofacial defects, vision and hearing defects, cardiac anomalies, and growth delays. CHARGE syndrome is caused by mutations in CHD7, encoding a ATP-dependent chromatin remodeling protein that regulates downstream target gene expression. Here we tested, using heterozygous loss of function Chd7 mutant mice (Chd7+/−), whether Chd7 deficiency affects retinoic acid signaling and if altered retinoic acid signaling in utero influences Chd7 mutant phenotypes. We focused specifically on the inner ear, due to the highly penetrant semicircular canal defects observed in both humans and mice with CHD7 deficiency. RNA-seq of microdissected e10.5 inner ear tissues revealed 50% or greater upregulation of 425 genes and downregulation of 146 genes in Chd7−/− compared with wild type. Several genes emerged as potential targets of CHD7, including genes encoding retinoic acid synthetase enzymes (Alldh), degradation enzymes (Cyp26), and receptors (Rar, Rxr). To test whether retinoic acid signaling during pregnancy influences Chd7 deficiency phenotypes, we administered retinoic acid or vehicle to pregnant mice at e17.5. Chd7−/− and wild type embryos were dissected at e14.5 and inner ear structures analyzed after paint-filling. Retinoic acid treatment worsened, or had no effect on, Chd7−/− inner ear defects, whereas vehicle treatment partially corrected the typically fully penetrant semicircular canal dysgenesis. Levels of retinoic acid were monitored in Chd7 mutant vs. wild type embryos using RARE (retinoic acid response element) reporter mice, which express β-galactosidase in response to retinoic acid. Chd7−/− embryos exhibited mildly increased β-galactosidase expression in the cochleovestibular ganglion, consistent with enhanced RA signaling. Aberrant retinoic acid signaling and gene expression loss with loss of Chd7, together with partial rescue of Chd7 deficient phenotypes upon retinoic acid inhibition support an essential role for CHD7 in determining of inner-ear tissue specific phenotypes. These studies have implications for understanding the pathogenesis of phenotypes in CHARGE syndrome, including variable expressivity and reduced penetration of organ specific defects, and for enhancing ongoing efforts to develop effective therapies for CHARGE individuals.

2266T
Mhyre syndrome-causing SMAD4 mutations result in disorganization of extracellular matrix that is corrected by losartan treatment. P. Pickens1, S. Mitthaler2, O. Sabatino3, J. John Tolmie2, D. Melo2, M.C. Schiaffino4, M. Filacamo5, G. Andria6, N. Brunetti-Pierri7,3. 1) Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Ferguson-Smith Department of Clinical Genetics, Yorkhill Hospital, Glasgow, UK; 3) Department of Translational Medicine, Federico II University of Naples, Naples, Italy; 4) Clinica Pediatrica, Istituto G. Gaslini, Genova, Italy; 5) Centro Diagnostico Genetica e Biochimica delle Malattie Metaboliche, Istituto G. Gaslini, Genova, Italy.
Mhyre syndrome (MIM 139210) is an autosomal dominant connective tissue disorder that presents with short stature, short hands and feet, facial dysmorphic features, muscle hypertrophy, thickened skin, and deafness. Mhyre syndrome is caused by recurrent missense mutations affecting codon ile500 of SMAD4 encoding a transducer mediating transforming growth factor (TGF-β) signaling. We investigated the functional consequences of SMAD4 mutations in Mhyre syndrome fibroblasts and observed altered expression of genes encoding matrix metalloproteinases and related inhibitors and a defect of extracellular matrix deposition. Disruption of microfibril network results in increased TGF-β bioavailability and increased TGF-β signaling in Mhyre syndrome fibroblasts. Mhyre syndrome commonly affects the connective tissue disorder, can be prevented by losartan, a TGF-β antagonists and angiotensin II type 1 receptor (AT1) blocker. In this study, we investigated whether losartan is effective at improving the extracellular matrix deposition defect of Mhyre syndrome cells. We showed that losartan normalizes SMAD2 phosphorylation, restores balance of metalloproteinases and related inhibitors, and improves the extracellular matrix deposition in fibroblasts from Mhyre syndrome patients. The results of this study may pave the way towards therapeutic applications of losartan in Mhyre syndrome.
2268F Tumor associated macrophages in neurofibromatosis. C. Prada1,2, E. Jousma3, T. Rizvi2, J. Wu4, D. Dunn5, N. Ratner2. 1) Dept Pediatrics Genetics, Cincinnati Children’s Hosp Med Ctr, Cincinnati, OH; 2) Divisions of Experimental Hematology and Cancer Biology, Cincinnati Children's Hosp Med Ctr, Cincinnati, OH; 3) Center for Genomic Medicine and Metabolism, Cardiovascular Foundation of Colombia, Floridanablanca, Colombia. Background: Plexiform neurofibromas (PNFs) are one of the most common and debilitating complications of Neurofibromatosis type 1 (NF1). In animal models, NF1-/- Schwann cells secrete large amounts of chemotractants (RANTES, SCF, M-CSF, and VEGF), enhancing migration of mast cells toward the peripheral nerves. Macrophages can be recruited by several cytokines including RANTES and VEGF. We previously identified macrophages infiltrating mouse PNFs. Macrophage can have anti-tumor (M1) or pro-tumor (M2) role at different cancer stages. We hypothesized that macrophage infiltration correlates with tumor progression and macrophage depletion induces cell death in Schwann cells. Methods: We tested this hypothesis in the Dhh-Cre; Nf1flox/flox mouse model of neurofibroma and human PNFs and MPNSTs. Immunohistochemistry was performed to quantify macrophage infiltration and macrophage subtype (M1 and M2 markers) in peripheral nerves and tumors at different disease stages to evaluate if increasing numbers of iba1+CD11b+ cells correlates with progression (n= 20). Then, we treated 7 Dhh-Cre; Nf1flox/flox mice with 50mg/kg of minocycline, a broad-spectrum antibiotic that partially depletes macrophages, by i.p. injection and 7 Dhh-Cre; Nf1flox/flox mice with vehicle control, both for 1 month, to evaluate if macrophage depletion leads to cell death in tumors. Results: Large numbers of macrophages stained with iba1 were seen in the peripheral nerve and PNFs from Dhh-Cre; Nf1flox/flox mouse model. Minocycline treatment decreased macrophage numbers. Most commonly, the majority of iba1+CD11b+ cells in the cell-rich and cell-poor regions of PNFs. Flow sorting from human neurofibromas showed an average of 8% (range 1.5 to 14%) of CD11b+ cells in the PNF. Macrophage treated tumors decreased iba1+ macrophages and increased Schwann cell death supporting a role for macrophages on Schwann cell survival in PNFs. Conclusion: Macrophages are a major component of NF1 peripheral nerve tumors in human and mouse with a direct effect on tumor progression and may have applications for the treatment of other human diseases. You may contact the first author (during and after the meeting) at fozsolak@ranarx.com.

2269T OLGIOThERAPEUtic STRATEGIES FOR THE TREATMENT oF FRIEDRICH’S atAXIA. O. Ozols8, D. Jun Li, D. Parekh, D. Knowlton, M. Wyss9, R. Subramanian, J. Barsoum. RaTherapeutics, Cambridge, MA. Friedreich’s ataxia (FRDA) is a recessively inherited disorder that arises due to cellular depletion of frataxin (FXN) protein resulting in mitochondrial dysfunction. FRDA is a progressive neuromuscular disease which lacks any FDA-approved therapy. The protein coding sequence of FXN is normal in the majority of FRDA patients, and the causative basis of this disease is the under-expression of the FXN gene, suggesting that upregulation of endogenous FXN expression could be an effective therapy. The most common molecular cause of this currently incurable disease is the expansion of GAA/TTC triplet repeats in the first intron of FXN gene. Repeat expansion beyond a certain threshold causes transcriptional defects which reduce FXN mRNA and protein levels. Despite long-standing research in the pathogenesis of FRDA, the means by which GAA-repeat number elevation leads to transcriptional silencing is not clear. Unusual DNA-DNA and RNA-RNA interactions formed in the long triplet repeat region leads to formation of a heterochromatin-like structure are among the hypotheses being considered. In order to gain clues into the mechanisms responsible for the FXN deficit in FRDA, we undertook genome-wide analyses to examine the global and local RNA structure and chromatin structure and composition changes in FRDA patient cells. Epigenetic screens identified two chromatin modifying complexes as being important in establishing and/or maintaining repeat expansion-induced transcriptional repression at the FXN locus. We identified a novel non-coding RNA (ncRNA) potentially responsible for directing the localized epigenetic silencing of the FXN gene. Degrading this ncRNA led to at least partial heterochromatin reversal and FXN mRNA and protein level upregulation to therapeutically significant levels. The oligonucleotide-based therapeutic strategy was shown to be well-tolerated in human tissue, indicating this approach is a promising candidate for the treatment of FRDA and may have applications for the treatment of other human diseases. You may contact the first author (during and after the meeting) at fozsolak@ranarx.com.

2270F Positive Effects of Short Course Androgen Therapy on the Neurodevelopmental Outcome in Boys with 47, XY Syndrome at 9 Years of Age. C.A. Sarmiento-Sprouse1, E. Stapleton1, C. Sprung2, T. Sadik3, F.L. Mitchell3, A.L. Gropman1. 1) George Washington University, Washington, DC; 2) Department of Neurology, Children’s National Medical Center, Washington, DC; 3) Neurodevelopmental Diagnostic Center for Young Children, Davidsonville, MD. Background: Positive effects of early androgen treatment on neurodevelopmental performance in prepubertal males with 47, XY have been documented at 36 and 72 months, giving support to the link between neurobiological development and androgens. Furthermore, our previous study demonstrated improved weight gain and catch-up growth, consistent with early androgen treatment. Purpose: The aim is to determine if an early course of androgen treatment (3 mg/kg of testosterone enanthate, 25mg, each) could have a positive impact on OXY boys at 9 years of age. Methods: 59 prenatally diagnosed males with karyotypes of 47, XY participated with one group (n=22) receiving androgen treatment in infancy and the second group untreated (n=37). Treatment of other human diseases. You may contact the first author (during and after the meeting) at fozsolak@ranarx.com.

Results: There was a significant positive treatment effect in multiple visual motor domains (VP P=0.0012, MC P=0.0129, VMI P=0.0151). A positive treatment effect was observed on the BOT (Manual Coordination P=0.0003, Bilateral Coordination P=0.0001, Body Coordination P=0.0038, Speed/Agility P=0.0328, Strength P=0.0161 Upper Limb Coordination P=0.0004, and Strength/Agility P=0.0134). Discussion: Long-term improved function has been observed in neurodevelopmental performance in XY males at 36 and 72 months and now at 108 months after prenatal androgen replacement. Macrocephaly is generally a low-risk procedure in humans, this route of delivery may be suitable for drug targeting with reduced drug exposure of the mother and fetus. The complete spectrum of hair, sweat gland and dentition response following prenatal EDA-A1 replacement via maternal injection corrected the developmental performance in prepubertal males with 47, XXY more than 7 years ago, giving support to the link between neurobiological and genetic variables. Conclusions: Prenatal therapy in developmental disorders: drug targeting via intra-amniotic injection (IAI) to treat X-linked hydropptic ectodermal dysplasia. K. Hermes1, P. Schneider2, K. Kreig2, A. Dang2, K. Huttner4, H. Schneider4. 1) German Competence Centre for Children with Ectodermal Dysplasias, Department of Pediatrics, University of Erlangen-Nürnberg, Germany; 2) Department of Biochemistry, University of Lausanne, Switzerland; 3) German Cancer Research Center, Heidelberg, Germany; 4) Edimer Pharmaceuticals, Inc., Cambridge, USA. Background: X-linked hydropthic ectodermal dysplasia (XLHED), the most common inherited disorder of ectoderm development, arises by a lack of the signaling molecule EDA-A1. In the Tabby XLHED mouse model, prenatal EDA-A1 replacement via maternal injection corrected the developmental abnormalities to a far greater extent than postnatal administration in newborns. This suggests that achieving physiological levels of therapeutically effective levels of corrective protein in the human fetus, and additionally exposes the mother to high serum levels of the exogenous molecule. We hypothesized that direct injection of EDA-A1 replacement protein into the amniotic sac of Tabby (EDA-A1-deficient) mice via intra-amniotic administration would result in sustained drug exposure at levels sufficient for successful treatment of XLHED. Methods: ED1200, a human IgG1:EDA-A1 fusion protein, was tested for stability in AF using a binding ELISA. Subsequently, ED1200 was injected into amniotic sacs of pregnant wild-type mice to evaluate drug uptake and pharmacokinetics. Fetal and maternal serum levels were monitored. Based on these results, ED1200 at a dose of 100 mg/kg fetal weight was administered intra-amniotically to E15 Tabby mouse fetuses and was shown to be well-tolerated. Intra-amniotic administration to E15 wild-type mice (35 mcg ED1200/fetus) resulted in substantial fetal uptake with mean serum levels of 9.0 mcg/ml and 1.2 mcg/ml at 6 hours and 96 hours, respectively. Maternal serum levels remained <0.1 mcg/ml and were not be detected by ELISA. ED1200 was administered via intra-amniotic injection to E15 proved to correct the XLHED phenotype in offspring only partially. The complete spectrum of hair, sweat gland and dentition response following E15 intra-amniotic administration of ED1200 is being evaluated through educational and social intervention. Conclusions: Intra-amniotic protein application to mice may lead to rapid fetal uptake, sustained substantial serum levels, and efficacy comparable with intravenous injection. As amniocentesis is generally a low-risk procedure in humans, this route of delivery may be suitable for drug targeting with reduced drug exposure of the mother and longer bioavailability due to a reservoir function of AF. It may, thus, represent a novel paradigm for treatment of disorders in early human development.
A Novel, Selective and Orally-available Glucosylceramide Synthase Inhibitor for Substrate Reduction Therapy of Fabry Disease. J. Mar-shall1, K. Ash1, E. Budman1, B. Rangarajan1, J. Nieuwsgeld1, R.J. Desnick1, R.K. Scheule1, J.P. Leonard1, S.H. Cheng1, 1) Genzyme, Framingham, MA; 2) Mount Sinai School of Medicine, New York, NY.

Fabry disease, an X-linked glycosphingolipid storage disorder, is caused by a deficiency of α-galactosidase A (α-gal). Resulting progressive accumulation of globotriaosylceramide (GL-3) and lyso-GL-3 leads to kidney, heart, and cerebrovascular disease. Presently, Fabry disease is managed by periodic infusions of recombinant α-gal (enzyme-replacement therapy; ERT). However, the inability of ERT to completely address disease manifestations in the heart and kidney has encouraged the development of alternative therapies. We had previously demonstrated that substrate reduction therapy (SRT) through antagonism of glucosylceramide synthase (GCS) can delay the accumulation of GL-3 in a mouse model of Fabry disease. Here, we describe the merits of a novel GCS inhibitor (Genz-682452) with favorable pharmacological properties and safety profile in Fabry mice. Treatment of Fabry mice with Genz-682452 starting at 3 months of age resulted in greater correction of a variety of disease biomarkers than when treatment was initiated in 12 months-old mice. Mice administered Genz-682452 exhibited significantly lower tissue levels of GL-3 and lyso-GL-3 and a delayed progression of a thermal nociceptive response than their untreated counterparts. SRT with Genz-682452, perhaps because it has a different biodistribution profile, was more effective than ERT at reducing the levels of the glycosphingolipids in the kidney, heart and CNS, organs that were not well-served by ERT. Importantly, mice treated by both ERT and SRT showed the greatest response suggesting the therapies are both complementary and additive. These results affirm the potential of SRT as an alternative and potentially adjuvant therapy for Fabry disease.

Successful pregnancy and lactation in a woman with mucopolysaccharidosis type I treated with laronidase. Y. Xue1, M. Castorina1, D. Antuzzi1, C. Sung1, S. Richards1, G. Cox1, 1) Genzyme, a Sanofi company, Cambridge, MA, USA; 2) Dipartimento di Tutela della Donna e della vita Nascente, Pediatría; Università Cattolica del Sacro Cuore, Rome, Italy; 3) Laboratorio di Neonatologia, Dipartimento di Tutela della Donna e della vita Nascente, Pediatría; Università Cattolica del Sacro Cuore, Rome, Italy.

Background: Few pregnancies have been reported in women with MPS I, a lysosomal storage disorder caused by α-L-iduronidase deficiency and subsequent accumulation of glycosaminoglycans (GAG) throughout the body. With the availability of laronidase enzyme replacement therapy (Aludrase®, Genzyme, a Sanofi company, Cambridge MA, USA) more women with MPS I may become pregnant and want to breastfeed their infants. Although in animals, laronidase has had no effect on fertility or pregnancy, the lack of human data has led to the recommendation that laronidase be used in pregnancy only if needed and to be used with caution during breastfeeding. Methods: We describe an ongoing, prospective, open-label trial funded by Genzyme (ALID 01803, NCT00418821) of women with MPS I who plan to receive laronidase during pregnancy and while breastfeeding. Following local ethics committee approval and signed informed consent, the first patient has completed the study and the results are presented below. The effects of laronidase on the mother during pregnancy and on her infant for 12 months after delivery were assessed periodically through physical examinations, IgG and IgM anti-laronidase antibody titers, urinary GAG, laronidase activity in breast milk, and developmental testing in the infant. Results: The patient, diagnosed with MPS I Scheule phenotype, was treated with laronidase for 3 years, enrolled in the trial during her second pregnancy and continued to receive laronidase throughout her pregnancy and while breastfeeding. A healthy 2.5 kg boy was delivered by elective cesarean section at 37 weeks gestation. The baby was breastfed for 3 months. Laronidase was not detected in breast milk. Anti-laronidase IgG antibodies were present in the mother and in the umbilical cord blood at birth. However, the infant had no detectable anti-laronidase IgM antibodies, and the IgG titer declined over time, consistent with passive maternal transmission of antibodies. Urinary GAG levels in the infant were always normal. The infant was healthy and had normal development through 12 months. No drug-related adverse events occurred. Conclusions: We report the first known case of a successful pregnancy outcome and normal development in an infant whose mother continued laronidase treatment during pregnancy and lactation. Further data are necessary to confirm the safety of laronidase during pregnancy and breastfeeding.

ENCORE: a randomized, controlled, open-label non-inferiority study comparing eliglustat to imiglucerase in Gaucher disease type 1 disorder or enzyme replacement therapy (ERT) in patients with and without liver enlargement. M. Balwani1, T.M. Cox2, G. Drellichman2, R. CRAVO2, T. Burrows2, A.M. Martins2, E. Lukina2, B. Rosenblom2, L. Ross2, J. Angell2, A.C. Puga2, 1) Dept Human Gen, Mt Sinai Med Ctr-New York, New York, NY, USA; 2) University of Cambridge, Department of Medicine, Addenbrooke’s Hospital, Cambridge, UK; 3) Hospital de Niños Ricardo Gutierrez, Buenos Aires, Argentina; 4) HEMORIO, Rio de Janeiro, RJ, Brasil; 5) Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA; 6) Universidade Federal de São Paulo, São Paulo, SP, Brasil; 7) Hematology Oncology Medical Center, Moscow, Russia; 8) Tower Hematology Oncology Medical Group, Beverly Hills, CA, USA; 9) Genzyme, a Sanofi company, Cambridge MA, USA.

Background: Gaucher disease type 1 is a multi-systemic lysosomal storage disorder caused by a deficiency of α-glucosidase (a glucosidosis type I treated with laronidase).

Here, we describe the merits of a novel GCS inhibitor (Genz-682452) with favorable pharmacological properties and safety profile in Fabry mice. Treatment of Fabry mice with Genz-682452 starting at 3 months of age resulted in greater correction of a variety of disease biomarkers than when treatment was initiated in 12 months-old mice. Mice administered Genz-682452 exhibited significantly lower tissue levels of GL-3 and lyso-GL-3 and a delayed progression of a thermal nociceptive response than their untreated counterparts. SRT with Genz-682452, perhaps because it has a different biodistribution profile, was more effective than ERT at reducing the levels of the glycosphingolipids in the kidney, heart and CNS, organs that were not well-served by ERT. Importantly, mice treated by both ERT and SRT showed the greatest response suggesting the therapies are both complementary and additive. These results affirm the potential of SRT as an alternative and potentially adjuvant therapy for Fabry disease.

ENCORE: a randomized, controlled, open-label non-inferiority study comparing eliglustat to imiglucerase in Gaucher disease type 1 disorder or enzyme replacement therapy (ERT) in patients with and without liver enlargement. M. Balwani1, T.M. Cox2, G. Drellichman2, R. CRAVO2, T. Burrows2, A.M. Martins2, E. Lukina2, B. Rosenblom2, L. Ross2, J. Angell2, A.C. Puga2, 1) Dept Human Gen, Mt Sinai Med Ctr-New York, New York, NY, USA; 2) University of Cambridge, Department of Medicine, Addenbrooke’s Hospital, Cambridge, UK; 3) Hospital de Niños Ricardo Gutierrez, Buenos Aires, Argentina; 4) HEMORIO, Rio de Janeiro, RJ, Brasil; 5) Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA; 6) Universidade Federal de São Paulo, São Paulo, SP, Brasil; 7) Hematology Oncology Medical Center, Moscow, Russia; 8) Tower Hematology Oncology Medical Group, Beverly Hills, CA, USA; 9) Genzyme, a Sanofi company, Cambridge MA, USA.

Background: Gaucher disease type 1 is a multi-systemic lysosomal storage disorder caused by a deficiency of α-glucosidase (a glucosidosis type I treated with laronidase).
ENGAGE: A Phase 3, randomized, double-blind, placebo-controlled, multi-center study to investigate the efficacy and safety of eliglustat in adults with Gaucher disease type 1 (GD1): 9 month results.

S. Packman1, D. Amato2, M. Dasouki3, G. Pastores4, S. Assouline5, M. Balwani6, P.K. Mistry7, S. Shankar8, M.J. Peterschmitt9, 1) Dept Peds/Div Med Gen, Univ California Sch Med, San Francisco, CA; 2) Mount Sinai Hospital, Toronto, Canada; 3) University of Kansas Hospital, Kansas City, KS, USA; 4) New York University School of Medicine, New York, NY, USA; 5) Hebrew General Hospital, Montreal, Quebec, Canada; 6) Mt. Sinai Hospital, New York, NY, USA; 7) Yale University School of Medicine, New Haven, CT, USA; 8) University, Atlantis, GA, USA; 9) Genzyme, a Sanofi company, Cambridge MA, USA.

Background: GD1 is one of the most common lysosomal storage disorders. Deficiency of the lysosomal enzyme glucocerebrosidase leads to accumulation of glucocerebroside in the spleen, liver, and bone marrow. Eliglustat, a novel oral substrate reduction therapy that selectively inhibits glucocerebrosidase, is in development for the treatment of GD1. ENGAGE (NCT00691202) is a randomized, double-blind, placebo-controlled, Phase 3 trial sponsored by Genzyme, a Sanofi company investigating the efficacy and safety of eliglustat in untreated adults with GD1. Methods: Forty patients (mean age: 31.8 years; 20 males) with splenomegaly and thrombocytopenia and/or anemia were stratified by spleen volume and randomized 1:1 to receive eliglustat (50 or 100 mg BID depending on plasma levels) or placebo for 9 months. The primary efficacy endpoint was percent change in spleen volume (multiples of normal). Other efficacy measures included hemoglobin, liver volume, and platelets. Bone endpoints included bone marrow burden (BMB) scores and bone mineral density changes (DXA). Quality of life assessments included the Gaucher Disease Specific QoL Questionnaire (DSQ) and the Gaucher DS3. Safety monitoring included adverse event reporting, and lab and ECG evaluations. Results: In patients receiving eliglustat vs. placebo, mean spleen volume decreased (-28% vs. +2%, P<0.0001), mean hemoglobin increased (0.9 vs. -0.5 g/dL, P<0.0006), liver volume decreased (-5.2% vs. +1.4%, P=0.0072), and platelets increased (+32% vs. -9.06%, P=0.0011). Significant improvements (eliglustat vs. placebo) were observed for total (-1.1 vs. 0.0, P=0.002), spine (-0.6 vs. 0.1, P=0.002), and femur (-0.5 vs. 0.0, P=0.026) BMB scores. Although patients with symptomatic bone disease were excluded, absolute change in total spine DXA T-scores approached significance (LS mean treatment difference=-0.2, P=0.06). Burden of disease, as measured by the Gaucher DS3, was significantly reduced following treatment (LS mean treatment difference=0.2, P=0.06). Quality of life improvements (eliglustat vs. placebo) were observed for total (-1.1 vs. 0.0, P=0.002), spine (-0.6 vs. 0.1, P=0.002), and femur (-0.5 vs. 0.0, P=0.026) BMB scores. Although patients with symptomatic bone disease were excluded, absolute change in total spine DXA T-scores approached significance (LS mean treatment difference=-0.2, P=0.06). No patient discontinued due to an AE, all of which were classified as mild to moderate; 39/40 patients transitioned into the ongoing open-label trial. Arthralgia and nasopharyngitis occurred in >10% of eliglustat vs. placebo patients. Conclusion: ENGAGE met its primary and secondary efficacy endpoints. Significant effects on bone marrow and a trend toward BMD improvement in spine were observed. Eliglustat was generally safe and well-tolerated.
2280F
Drug screening using the dystroglycan null zebrafish. G. Kawahara1, J. Widrick2, V. Gupta3, A. Myer1, M. Gasperini4, A. Beggs1, L. Kunkel1, 2, 3. 1) Division of Genetics, Program in Genomics, The Manton Center for Orphan Disease Research, Boston Children’s Hospital, Boston, MA, USA; 2) Department of Pediatrics, Harvard Medical School, Boston, MA, USA; 3) Department of Genetics, Harvard Medical School, Boston, MA, USA.
Zebrafish are an ideal model for biomedical research, as zebrafish genetic models mimic much of the pathology of human diseases. Dystroglycan null fish (patchytail, dag1c1500) are an excellent model of the human dystroglycanopathies as they show disturbed muscle structure and a severe reduction of birefringence at 3-7 days post fertilization (dpf). The birefringence is a result of dorsal skeletal muscle deterioration and weakness that ultimately results in lethality of most dystroglycan null mutants at 20 dpf. We have performed a preliminary drug screen of 1120-chemicals from a small, commercial molecular library, Prestwick collection using homozygous patchytail zebrafish to identify potential therapeutic chemicals for treating the dystroglycanopathies. Embryos from heterozygous dag1c1500/wt zebrafish matings were cultured in normal fish water containing pools of up to 8 compounds from the Prestwick collection, and were incubated from 1 dpf to 4 dpf. At 4 dpf the treated fish were examined by birefringence to determine if there was any reduction from the expected 25% fry showing abnormal birefringence. Fish treated with some of the pooled chemicals were genotypically confirmed to be dystroglycan null by birefringence compared to the untreated control. This suggests that we may have successfully identified efficacious compounds for prevention of the phenotype caused by the dystroglycan gene mutation. In this short-term screen of 1120 compounds, we found eleven candidate drugs that prevent muscle pathology by significantly reducing the proportion of fry with abnormal birefringence at 4 days. Interestingly, five of the candidate drugs are involved in gamma-aminobutyric acid regulation. Dystroglycan null embryos were treated with each of these candidate drugs for a long-term period, from 5 dpf to 20 dpf. This treatment indicated that one of the candidate compounds greatly increased survival of dystroglycan null fish. Analysis of these treated surviving dystroglycan null fish showed they had normal muscle structure and normal neuromuscular junctions, whereas these structures are observed as abnormal absent or fragmented in untreated fish. A birefringence screen will likely be informative for development of drugs to treat human dystroglycanopathies. The discovery of a small molecule and specific therapeutic pathway that might mitigate progression of this disease is highly relevant and significant.

2282F
Sildenafil citrate results in upregulation of heme oxygenase 1 and alleviation of symptoms in the mdx mouse model of Duchenne muscular dystrophy. M. Gasperini1, G. Kawahara1, 2, J. Widrick1, M. Alexander1, 2, L. Kunkel1, 2, 3. 1) Division of Genetics, Program in Genomics, The Manton Center for Orphan Disease Research, Boston Children’s Hospital, Boston, MA, USA; 2) Department of Pediatrics, Harvard Medical School, Boston, MA, USA; 3) Department of Genetics, Harvard Medical School, Boston, MA, USA.
Duchenne muscular dystrophy (DMD), the most common form of muscular dystrophy, is caused by lack of the dystrophin protein, with no effective treatment at present. Zebrafish are a powerful tool for high throughput drug screening in vivo. From therapeutic drug screens using dystrophin-deficient sagie zebrafish, we have identified four drugs that increased affected fish survival rates at 20 days post fertilization. These drugs induced upregulation of heme oxygenase 1 (Hmx1) protein expression in surviving dystrophin null fish. One of these compounds, sildenafil citrate, is reported as a known effective mammalian drug for DMD. To confirm its therapeutic effects and its upregulation of Hmx1 in a DMD mouse model, sildenafil was administered to dystrophic null mdx5cv mice. Sildenafil was dissolved to a 400 mg/L concentration in sterile-filtered, acidified drinking water. Two male and two female 4 week-old mdx5cv mice were given ad libitum access to treated water for 6-8 weeks. During treatment, mice were tested for exercise tolerance by monitoring activity for 6 minutes immediately before and after 15 minutes of treadmill running. After the drug treatment period, skeletal muscle was histologically and physiologically examined to check for improvement of symptoms, as well as changes in Hmx1 expression. Whereas exercise tolerance at 2 and 4 weeks of drug administration showed no significant improvement, sildenafil treated mice displayed alleviated exercise fatigue symptoms after 6 weeks of drug administration. After treadmill running, their total seconds spent resting, centimeters traveled, and number of rearing events were significantly improved over the extremely fatigued vehicle-treated dystrophic mice. The sildenafil receiving mice also had a reduced number of centralized nuclei in skeletal muscle fibers in comparison to dystrophic controls, as counted in Haematoxylin and Eosin stains of diaphragm and tibia anterior muscle slices. Immunoblot revealed that sildenafil treated mice showed increased expression in skeletal muscle when compared to vehicle-treated wildtype and mdx5cv mice. These results point to Hmx1 as an effective therapeutic pathway in fish and mouse models of DMD. Hmx1 and drugs that impact it will likely be informative for development of therapeutics to treat humans with DMD.

2287T
Developing Small Molecule Inhibitors of p97/VCP Disease Mutants for Neurodegenerative Diseases. T.-F. Chou1, C.C. Weinhi2, R.J. Deshaies3. 1) Division of Medical Genetics, Department of Pediatrics, Harbor-UCLA Medical Center and Los Angeles Biomedical Research Institute, Torrance, CA 90502, USA; 2) Department of Neurology, Washington University School of Medicine, 660 South Euclid Avenue, Box 8111, St. Louis, MO 63110, USA; 3) Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.
p97 AAA (ATPase Associated with diverse cellular Activities; also called VCP [Valosin-Containing Protein] ATPase) serves as a central player in a complex network of intracellular proteolytic pathways that mediate protein homeostasis, cell cycle progression, immunity, autophagy and other cellular functions.

In this study, we examined the effects of p97 ATPase inhibition in a mouse model of Duchenne muscular dystrophy, the mdx mouse. p97 gene knockdown or inhibition led to the reduction of dystrophin in the quadriceps femoris muscle, strongly implicating p97 in the pathogenesis of muscular dystrophy. Furthermore, we have shown that targeting p97 is sufficient to rescue mdx mice from the lethality of the disease.

To develop small molecule inhibitors of p97, we employed a high-throughput screening approach using a panel of over 1000 compounds. We identified four compounds that significantly inhibited p97 ATPase activity in vitro. These compounds were then tested in a DMD mouse model, where they were shown to rescue dystrophin expression in skeletal muscle fibers.

Our findings suggest that targeting p97 may be a promising therapeutic strategy for the treatment of muscular dystrophy and other diseases associated with impaired protein homeostasis.
Cerebral cavernous malformations (CCMs) are vascular malformations in the brain characterized by grossly dilated capillaries and vascular leak and hemorrhage, leading to stroke, seizure and other neurological complications. The disease occurs in sporadic or inherited (autosomal dominant) forms, the latter caused by mutation in one of three genes (CCM 1, 2 or 3). The vascular lesions in the familial cases follow a two-hit mutational mechanism of pathogenesis due to somatic mutation of the wild-type copy of the relevant CCM gene. Based on this two-hit model, we crossed heterozygous murine Ccm gene knockout alleles into mutant backgrounds with elevated rates of somatic mutations (p53 or Msh2 null alleles), and have generated robust murine models of CCM disease that fully and faithfully recapitulate the clinical phenotype. In parallel, in vitro studies have shown that the CCM proteins localize to endothelial cell-cell junctions and their loss leads to junctional instability associated with activation of RhoA and its effector, Rho kinase (ROCK). As a proof of principle, we recently demonstrated that ROCK inhibition using the drug fasudil reduces lesion burden in our murine Ccm1 model. In the current study we investigated the effect of ROCK inhibition in other Ccm murine genotypes, including a murine Ccm3 model, the most penetrant and morbid of the inherited forms in both the mouse and human. In all genotypes, fasudil treated mice exhibit decreased prevalence of CCM lesions compared to placebo controls. Lesions in treated animals exhibit fewer and smaller cavities (ie. ‘immature’ lesions) and are less likely to rupture with hemorrhage. Coming full circle, we have also now found that sporadic CCM lesions in human (the most common form of CCM) also exhibit somatic mutations in the same genes found mutated in the inherited cases. These combined data suggests mutation (somatic or inherited) of the CCM genes can also result in all forms of CCM. Thus, ROCK inhibition may be a viable therapeutic option for all forms of CCM disease, including the most severe inherited form, CCM3, and possibly the more common, sporadic cases.

Development of a high-throughput screen for mRNA splicing modifiers of IKBKAP. F. Urbina, M. Niblatt, E. Morini, M. Salani, S. Slaugenhaupt. Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA.

Familial Dysautonomia (FD) is an autosomal recessive sensory and autonomic neuropathy that primarily affects the central and peripheral nervous system. All patients with FD have an intronic splice site mutation in the IKBKAP gene (IVS20+6T→G) that leads to aberrant splicing, with partial skipping of exon 20. This exon skipping is tissue specific and leads to a reduction of IKAP protein that is most pronounced in the nervous system. This reduction of IKAP leads to a number of symptoms including absence of overflow tears, gastrointestinal reflux, scoliosis, and orthostatic hypotension. Previously, we showed that the plant cytokinin kinetin dramatically improves aberrant splicing in FD patient lymphoblast and fibroblast cells. Further, kinetin works to modulate splicing in vivo improving exon 20 inclusion and increasing the amount of IKAP protein in all tissues, including brain, in a mouse model. In order to work towards optimizing kinetin and identifying new compounds that might be useful as human therapeutics, we established both primary and secondary screening assays. Using a previously created FD minigene construct, a dual luciferase-based primary assay was developed and optimized for high-throughput screening of compounds. A luciferase inhibition counterscreen was also developed. A secondary qPCR assay to evaluate the level of normal IKBKAP mRNA was established to both evaluate luciferase inhibitors and “hit” compounds in both the primary assay and in patient cell lines. This screening funnel will allow for a large number of compounds to be evaluated in both a primary and patient-based secondary screen for splicing modulation. Several kinetin-like compounds have been evaluated to date and prove that our system will permit the identification of potential new therapies for this mRNA splicing disorder.

Rho Kinase Inhibition Therapy for Cerebral Cavernous Malformations. D.A. Marchuk1, D.A. McDonald3, R. Shenkar2, C. Shi1, C.J. Gallione2, C. Austin2, A.G. Mikat1, I.A. Award1. 1) Molec Gen & Microbiol, Duke Univ Med Ctr, Durham, NC; 2) Neurovascular Surgery Program, University of Chicago Medicine and Biological Sciences, Chicago, IL.

Rho kinase (ROCK). As a proof of principle, we recently demonstrated that ROCK inhibition using the drug fasudil reduces lesion burden in our murine Ccm1 model. In the current study we investigated the effect of ROCK inhibition in other Ccm murine genotypes, including a murine Ccm3 model, the most penetrant and morbid of the inherited forms in both the mouse and human. In all genotypes, fasudil treated mice exhibit decreased prevalence of CCM lesions compared to placebo controls. Lesions in treated animals exhibit fewer and smaller cavities (ie. ‘immature’ lesions) and are less likely to rupture with hemorrhage. Coming full circle, we have also now found that sporadic CCM lesions in human (the most common form of CCM) also exhibit somatic mutations in the same genes found mutated in the inherited cases. These combined data suggests mutation (somatic or inherited) of the CCM genes can also result in all forms of CCM. Thus, ROCK inhibition may be a viable therapeutic option for all forms of CCM disease, including the most severe inherited form, CCM3, and possibly the more common, sporadic cases.
2288F
Fusion with Angiopep-2 to create proteins that cross the blood-brain barrier and are taken up into cells. J.E. Lachowicz, M. Demeneu, A. Regina, D. Bolvin, A. Larocque, J.-P. Castaigne. Angiochem, Montreal, Quebec, Canada.

Enzyme replacement therapy has been used successfully to treat peripheral symptoms of lysosomal storage disorders. However, treatment of the severe CNS symptoms associated with many LSDs has been impeded by the inability of enzymes to cross the blood-brain barrier (BBB). Receptor-mediated transcytosis is one strategy for allowing large molecules to cross the BBB. The low density lipoprotein-like 1 (LRP-1) receptor binds over 40 natural ligands for BBB transcytosis. Incorporation of Angiopep-2 (An2), a peptide that binds LRP-1, into small molecules, peptides, and biologics such as mAbs has been shown to increase brain permeability. The most advanced example is a conjugate of An2 and paclitaxel (ANG1005). In contrast to paclitaxel, ANG1005 is brain-penetrant and is not recognized by the P-gp efflux pump at the BBB. In addition, reduction in tumor size has been shown with ANG1005 both in preclinical models and in patients with primary and metastatic brain tumors. A visual demonstration of the An2 conjugate technology for proteins has been achieved by creating a recombinant human An2-GFP fusion protein. In vitro, native GFP is not readily taken up by cells. In contrast to native GFP, An2-GFP incubation results in observable green fluorescence within cultured fibroblasts. Mice that are systemically treated with An2-GFP via carotid artery or tail vein show similar brain fluorescence to vehicle-treated mice, which is limited to the capillary basement membranes. However, An2-GFP treatment results in abundant intracellular green fluorescence in brain. Ten minutes post injection, fluorescence is largely concentrated in astrocytic end feet around capillaries, while 60 minutes post injection, a widespread fluorescence distribution is observed in neurons in all brain regions examined. At the subcellular level, punctate fluorescence is localized to presumptive endosomes, indicating active cellular uptake of An2-GFP. This study demonstrates that an An2-protein conjugate crosses the BBB and is taken up by cells in the brain, both attributes being critical for enzyme replacement therapy for CNS indications. We have applied this approach to an enzyme associated with mucopolysaccharidosis to create a brain-penetrant enzyme for enzyme replacement therapy.

2289T

Gaucher Disease (GD) is the most prevalent lysosomal storage disorder, caused by the deficiency of the lysosomal enzyme glucocerebrosidase (GCD). The accumulation of excessive glucocerebrosides in lysosomal compartments of macrophages causes hepatosplenomegaly, anemia and thrombocytopenia as well as bone pain and fractures. Enzyme replacement therapy (ERT) with GCD has been successfully used for the clinical treatment of GD for several years. All three currently approved ERT drugs are administered by intravenous infusion every 2 weeks. Although all enzymes are safe and efficient, the intravenous administration remains a limitation which affects patients’ quality of life. Thus, orally administered ERT would have the definite advantage of the well-established therapy mechanism, without the limitation of the intravenous administration and will provide continuous enzyme secretion. Oral administration of proteins is one of the challenges of the biopharmaceuticals industry particularly due to early degradation of the proteins in the digestive tract. One of the approved ERTs, taliglucerase alfa, is expressed in carrot cells. The use of carrot cells as natural vehicle of the expressed enzyme in recombinant GCD is intended for oral delivery of the enzyme, due to the protection from degradation provided by the composition of the plant cell wall. Enzyme expressed in the ProCellEx system, prGCD, has the optimal glycosylation profile, with exposed mannose for efficient uptake in target cells. This study shows the feasibility of this approach. Non clinical studies in rats and pigs fed with carrot cells expressing human GCD at either edible or repeat administrations, showed elevated levels of active GCD in both plasma and target organs (liver and spleen). Carrot cells expressing prGCD were shown to be safe and well tolerated by the animals and caused no major adverse symptoms. Clinical trial with oral administration of carrot expressing prGCD is on-going in GD patients. This study will evaluate enzyme delivery which allows daily intake and slow, continuous delivery of the drug. Once the enzyme is released in the blood stream it is expected to be similar to the approved ERT for which there is established clinical safety and efficacy data.

2290F
Preliminary Findings Evaluating Safety and Efficacy of Recombinant Human N-Acetylgalactosamine-6-Sulfatase in Pediatric Patients Less Than 5 Years of Age with Mucopolysaccharidosis IV (Morquio A Syndrome, MPS IVA). C. Haller1, S.A. Jones2, P. Harmatz3, M. Bialer4, R. Parini5, K. Martin1, P. Farmer1, P. Slator1. 1) Clinical Science, BioMarin Pharmaceutical Inc., Novato, CA; 2) St. Mary’s Hospital, CMFT, University of Manchester, MAHSC, UK; 3) Children’s Hospital and Research Center, Oakland, CA; 4) North Shore LJU Health System, Manhasset, NY; 5) Az. Ospedaliera S. Gerardo, Monza, Italy.

Preliminary results after 26 weeks of treatment from an ongoing study evaluating safety and efficacy of recombinant human N-Acetylgalactosamine-6-sulfatase (rhGALNS) in 15 MPS IVA patients <5 years of age are reported. The mean (range) age at baseline (BL) was 3.1 (0.8-4.9) years. Based on physical examination findings at BL, the majority of subjects had abnormal musculoskeletal features (14/15, 93.3%), abnormal general appearance (10/15, 66.7%), abnormal finding for HEENT (10/15, 66.7%) and corneal clouding (7/15, 46.7%). Most subjects had normal cardiovascular (12/15, 80.0%), respiratory (15/15, 100.0%), gastrointestinal (14/15, 93.3%), genitourinary (15/15, 100.0%), and neurologic systems (14/15, 93.3%) on exam. Standing height/length (n=15) was severely affected in many subjects; 7(46.7%) at <3rd, 3(20.0%) at ≥3rd<10th, 2(13.3%) at ≥25th<50th, and 3(20.0%) at ≥50th percentiles. Normalized urine keratan sulfate (uKS) was elevated 3 fold above the mean for age matched control population with a mean (range) of 35.9 (18.8-56.5) µg/mg creatinine (n=15). After 26 weeks of weekly infusions with 2 mg/kg/day of rhGALNS, the most commonly reported adverse events (AEs) were vomiting in 12 (80.0%), pyrexia in 11 (73.3%), and cough in 8 (53.3%) subjects. The majority of AEs were classified as ‘mild’ (10/15, 66.7%) or ‘moderate’ (3/15, 20.0%). Four serious AEs were reported in one patient: tonsillar hypertrophy, skin infection, sepsis and hypersensitivity. No subjects discontinued treatment, and rhGALNS had a similar safety profile as seen in older children and adults. In 8 subjects with 26 weeks of data, rhGALNS led to a substantial decrease in uKS in 7/8 subjects. Median decrease was 30.5 (±15.40%) after 2 weeks that was sustained at -35.2 (-15.57)% at 26 weeks. These results demonstrate the drug’s pharmacodynamic effects. Mean height/length for age z-scores did not demonstrate significant change from BL (-1.8 SD) to Week 26. In contrast, in these 8 patients the rhGALNS treated group for comparison, data in the literature indicate that these children experience growth failure in early childhood. The accumulation of KS begins early in the life of MPS IVA subjects, and is progressive and life limiting. Early intervention with rhGALNS treatment is well tolerated; produces decrease in keratan sulfate storage, and may show substantial benefit in young pediatric patients.
2291T Survival Rates and Timing of Initiation of Treatment with Enzyme Replacement Therapy (ERT) Among Patients with Classic Infantile-Onset Pompe Disease Enrolled In the Pompe Registry. P.S. Kishnani1, S. Jones2, A. van der Ploeg3, E. Mangen4, B. Byrnes5, V. Lelld6, N. Leslie7, S. Shankar8, P. Tanpaiboon9, D.W. Stockton10, J.B. Hennemann11, Z. Deveci11, J. Kemp11, J. Keutler12, Y-H. Chien12, 1) Division of Medical Genetics, Department of Pediatrics, New York Medical College, Valhalla, NY, USA; 2) Weill Cornell Medical College, New York, NY, USA; 3) National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA; 4) Department of Pediatrics, University of Florida, Gainesville, FL, USA; 5) Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK; 6) Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA; 7) Emory University School of Medicine, Atlanta, GA, USA; 8) Children’s National Medical Center, Washington, DC, USA; 9) Wayne State University School of Medicine, Detroit, MI, USA; 11) Charite Universitaetsmedizin Berlin, Berlin, Germany; 12) Genzyme, a Sanofi company, Cambridge, MA, USA; 13) National Taiwan University Hospital, Taipei, Taiwan.

Background: Pompe disease presents as a clinical spectrum with variable severity, progression, and muscle involvement. Early clinical diagnosis and initiation of enzyme replacement therapy (ERT), before the development of severe, irreversible symptoms, are important, especially in very young patients with the most severe form of the disease. Aim: To report the results of a descriptive analysis of the chances of survival and ventilator-free survival and its relation to timing of initiation of ERT among patients with symptom onset ≤12 months with cardiomyopathy (classic infantile-onset Pompe disease) in the Pompe Registry, sponsored by Genzyme. Methods: All patients in the Registry with symptom onset ≤12 months with evidence of cardiomyopathy and a record of treatment with ERT were eligible for analysis. Kaplan-Meier curves were fitted, stratified by patients with a recorded first ERT infusion <3 months or ≥3 months of age. Results: A total of 140 patients with symptom onset ≤12 months with evidence of cardiomyopathy were eligible for analysis. Patients who received their first ERT <3 months of age (n=36) had lower 3-months, 12-months, and 36-months (85% vs 73%) and 36-months (81% vs 64%) of follow-up. Invasive ventilator-free survival rates also were better for patients with initiation of ERT <3 months of age (n=24) compared with patients who started ERT ≥3 months of age (n=69) at 12 months of follow-up (94% for ERT initiation at ≤ 3 months of age vs 91% for ERT initiation at > 3 months of age) and more noticeably at 24 months (85% vs 73%) and 36 months (81% vs 64%) of follow-up. Ventilator-free survival rates also were better for patients with initiation of ERT <3 months of age (n=104) at 12 months of follow-up (94% for ERT initiation at <3 months of age vs 91% for ERT initiation at ≥3 months of age) and more noticeably at 24 months (85% vs 73%) and 36 months (81% vs 64%) of follow-up. Results were similar at all timepoints when patients from Taiwan were excluded. Conclusions: Earlier initiation of treatment in classic-infantile onset Pompe disease patients may improve the chances of survival and lead to better clinical outcomes overall and improvement of symptoms for these young, very ill patients.


PRX-102, is a chemically modified, plant-cell expressed recombinant human alpha-GAL-A enzyme (pH-alpha-GAL-A) developed as enzyme replacement therapy (ERT) for the treatment of Fabry disease. There are currently two ERTs of recombinant human alpha-GAL-A commercially available: agalsidase-alfa and agalsidase-beta. The plant cell expressed recombinant human alpha-GAL-A has been modified by cross-linking and the plant cell expression system allows for the production of PEG-PCL nanocarriers, varying the production process and surfactant concentration of enzyme modification and protecting it from host immune system until the enzyme reaches the target organ. We hypothesize that PRX-102 is active and localizes to the intracellular target, the lysosome, in Fabry patients' cells. PRX-102 presents higher stability under physiologically relevant conditions, and extended circulation residence time as compared to available commercial ERT. Furthermore, pharmacodynamic studies using a mouse model of Fabry disease show higher uptake in target organs than available commercial ERT and significant reduction of Gb3 accumulation. The extended circulation time, together with the increased uptake of PRX-102 as compared to commercial ERT, has the potential for a more efficient ERT for the treatment of Fabry disease. Currently, the PRX 102 enzyme is under evaluation in Fabry patients as part of a clinical study.

2293T The Influence of a Polymorphism in the Gene Encoding Angiotensin Converting Enzyme (ACE) on Treatment Outcomes in Late-Onset Pompe Patients Receiving Alglucosidase Alfa. R.C. Baek, R.E. Palmer, R.J. Pomponio, A.J. Movie-Wylie, Genzyme, a Sanofi Company, Framingham, MA.

Angiotensin Converting Enzyme (ACE) catalyzes the conversion of angiotensin I to the vasoactive peptide angiotensin II, and degrades bradykinin, a potent vasodilator. One outcome of these two actions is vasconstriction and an increase in blood pressure. Half of the variation in human ACE activity can be accounted for by an insertion/deletion (I/D) allele in intron 16 of the ACE gene. An insertion of an Alu repeat in intron 16 results in lower ACE activity (I allele), and conversely, the deletion (D allele) results in higher ACE activity. Furthermore, the I allele has been reported to be associated with the predominance of slow-twitch muscle fibers (Type I) and the D allele with fast-twitch muscle fibers (Type II). Two publications recently suggested that the ACE I/D polymorphism modifies the clinical presentation of Pompe patients and influences treatment outcomes following alglucosidase alfa enzyme replacement therapy (ERT). Specifically, patients with the D/D genotype presented with an earlier onset of disease, higher CK levels at diagnosis, increased pain, and more severe disease progression (Fillippi et al, 2010), and had poorer treatment outcomes on ERT (Ravaglia et al, 2012). We investigated these findings in a large cohort of late-onset Pompe patients included in a randomized, placebo-controlled trial of alglucosidase alfa. Our results also suggest that patients carrying two ID alleles demonstrate an attenuated response to treatment relative to the I/D and I/ I genotypes. This result was not associated with the antibody response to alglucosidase alfa. An investigation into the disease status of patients at entry into the clinical trial is ongoing.

2294F Polymeric Nanocarriers as Vehicles for the Treatment of Lysosomal Storage Diseases. M. Latorre-Esteses, A. Roman. Chemical Engineering, University of Puerto Rico- Mayagüez Campus, Mayagüez, PR.

Lysosomal Storage Diseases (LSDs) are a group of inherited genetic diseases caused by mutant lysosomal enzymes, leading to the accumulation of undigested macromolecules in the lysosome, thus causing increases in lysosome size and number, cellular dysfunction, clinical abnormalities, and premature death. Some LSDs can be treated with Enzyme Replacement Therapy (ERT) through intravenous administration of a recombinant enzyme in replacement of the defective enzyme. However, this is an expensive and inefficient method with adverse side effects associated with the high enzyme amounts required for the treatment, the need of post-translational modification of the enzyme and the host immune system response. We hypothesize that nanocarriers composed of Polyethylene glycol and Polycaprolactone (PEG-PCL) block copolymers can enhance ERT by eliminating the need of enzyme modification and protecting it from host immune system until lysosomal target is reached. We have designed these nanocarriers to remain stable at physiological pH and destabilize at acidic pH, allowing them to reach the target intact and degrade once inside the acidic lysosome, only then releasing therapeutic cargo into affected cellular organelle. In order to obtain the appropriate nanocarriers for ERT application, we synthesized a group of PEG-PCL nanocarriers, varying the production process and surfactant concentration. The most suitable combination was found by performing dynamic light scattering analysis (DLS), gel permeation chromatography (GPC), Thermogravimetric Analysis (TGA) and several spectroscopic techniques. We found the nanocarriers are not cytotoxic in cell lines tested. We are currently working on the reproducibility of syntheses and protein loading capacity into nanocarriers.
2295T

Mutations of GBA1, the gene encoding glucocerebrosidase, represent a common genetic risk factor for developing the synucleinopathies Parkinson’s disease (PD) and dementia with Lewy bodies (DLB). PD patients with or without GBA1 mutations also exhibit lower enzymatic levels of glucocerebrosidase in the central nervous system (CNS), suggesting a possible link between the enzyme and the development of the disease. To probe this link further, we evaluated the efficacy of augmenting glucocerebrosidase activity in the CNS of a mouse model of Gaucher-related synucleinopathy (Gbα1D409V/D409V) and a transgenic mouse overexpressing A53T α-synuclein. Adeno-associated virus-mediated expression of glucocerebrosidase in the CNS of symptomatic Gbα1D409V/D409V mice completely corrected the aberrant accumulation of the toxic lipid glucosylsphingosine and reduced the levels of ubiquitin and protease-K-resistant α-synuclein aggregates. Importantly, hippocampal expression of glucocerebrosidase in Gbα1D409V/D409V mice (starting at 4 or 12 months old) also reversed their cognitive impairment when examined using the novel object recognition test. Overexpression of glucocerebrosidase in the CNS of A53T α-synuclein mice reduced the levels of soluble α-synuclein, suggesting that this glycosidase can modulate the development of α-synucleinopathies. Hence, increasing glucocerebrosidase activity in the CNS represents a potential therapeutic strategy for GBA1-related and non-GBA1-associated synucleinopathies.

2296F
Permanent genetic modification of dividing cells using episomally maintained S/MAR DNA vectors and the correction of a cancer phenotype in renal tumour cells. R. Hartbottle1,2, SP. Wong2. 1) DNA Vector Research Group, German Cancer Research Centre (DKFZ), Heidelberg, Germany; 2) Gene Therapy Research Group, Imperial College London, UK.

The simple, stable and efficient application of episomal DNA vectors to genetically modify dividing cells without the risk of integration-mediated genotoxicity provides a valuable tool in cell biology research. Here, we demonstrate the utility of Scaffold/Matrix Attachment Region (S/MAR) DNA plasmid vectors rapidly and simply generate novel genetically modified cell lines. In this study we utilize these vectors to model the restoration of a functional wild-type copy of the gene implicated in the renal cancer Birt-Hogg-Dube (BHD) in a cell-line (UOK257) derived from a BHD patient. Inactivation of the BHD gene, encoding a protein called folliculin (FLCN) has been shown to be involved in the development of sporadic renal neoplasia in BHD. Persistent genetic correction of UOK257 cells with an S/MAR-FLCN plasmid (UOK257-FS) restores FLCN expression and normalizes downstream TGFβ signalling. We demonstrate that UOK257-FS cells show a reduced growth rate in vitro and suppressed xenograft tumour development in vivo, compared to the original FLCN-null UOK257 cell line. We also show that mTOR signaling in serum-starved FLCN-restored cells is differentially regulated compared to the FLCN-deficient cells indicating the complex role of FLCN in separate signalling pathways. The novel UOK257-FS cell line will be a useful tool for studying different feedback loops in signalling pathways affected in BHD pathogenesis and provides further insight into the role of the FLCN in BHD. More significantly, this study demonstrates the suitability of episomally maintained S/MAR DNA vectors to successfully model the persistent functional expression of a therapeutic gene in a cancer cell line. It should also establish this class of vectors as effective tools for investigating signalling components differentially regulated in different cancers and to aid the identification of novel cancer markers for diagnosis and therapy. Additionally, this study illustrates the utility of this class of vectors for the genetic modification of cells without the risk of genotoxicity.

2297T

Menkes disease is a debilitating X-linked recessive neurometabolic disorder caused by mutations in a copper-transporting ATPase, ATP7A. No long-term, predictably effective treatment for affected individuals has been developed in the 50 years since the initial description of Menkes disease. While subcutaneous injections of copper may bypass the defect in intestinal copper absorption and normalize blood copper levels, at least a rudimentary level of ATP7A function is needed to restore proper intraneuronal copper transport, and attain normal neurodevelopment. Notably, early copper replacement in some affected patients whose mutations permitted as little as 5-10% residual ATP7A copper transport activity led to completely normal neurodevelopmental outcomes. We found previously that intracerebroventricular (ICV) administration of AAV (serotype 5) harboring an expression cassette with a compact version of human ATP7A had a salutary, synergistic effect with ICV copper administration in a lethal mouse model of Menkes disease, mo-br. This combination therapy led to higher brain copper concentrations, reduced neuropathology, and enhanced survival when administered to the lateral ventricles of newborn mo-br mice (Donsante et al. Mol Ther, 2011). A recently identified AAV serotype, rh10, shows broader neuronal tropism than AAV5, demonstrating robust transduction of multiple neuronal cell types, as well as choroid plexus epithelia. Broad neuronal transduction would be especially desirable in instances of a complete loss-of-function ATP7A mutation. Brain-directed rAAVrh10 represents a desirable route of administration from immunological and tissue target perspectives. Therefore, we have advanced our efforts in the mo-br model of Menkes disease by evaluating the efficacy of ICV administration of rAAVrh10:ATP7A in combination with subcutaneous copper histidine (CuHis). We demonstrated extended lifespan and brain neurochemical correction with this regimen. The administration of the treatment prior to days 16 to 18 postnatal resulted in complete rescue of the mo-br phenotype with ICV rAAVrh10 on day 2 with subsequent CuHis (10 μg) on day 5 (single administration of each). These dose-ranging preclinical studies presage application of this approach in human subjects with Menkes disease associated with severe loss-of-function ATP7A mutations.

2298F

Many studies have repeatedly demonstrated the safety of adeno-associated virus (AAV) vectors. However, one report has documented an increased incidence of hepatic cellular carcinoma (HCC) in neonatal mice after AAV gene therapy and further implicated insertional mutagenesis by the vector at the RIAN locus as a causative factor (Donsante et al., 2007). We have previously demonstrated the pre-clinical efficacy of AAV gene delivery using a neonatal lethal MMA murine model. Although the evaluation of genotoxicity was not the intended focus of our studies, the long-term surveillance of both AAV-treated MMA mice and control littermates has revealed that 75% (n = 48) of AAV8-CBA-Mut treated mice developed HCCs between 12 and 21 months compared to a 2% (n = 41) HCC rate in untreated littermates. In addition, 50% of the mice similarly treated with a control vector, AAV8-CBA-GFP, also developed HCCs, indicating that overexpression of the Mut transgene is not responsible for the development of HCC. We next performed a dose escalation study by increasing the AAV8 dose from 7 x 10¹⁰ GC/kg (n = 16) to 1 x 10¹¹ GC/kg (n = 19) and noted a corresponding increase in the incidence of HCC from 12% to 84%, respectively. As previously observed by Donsante et al., 2007, our preliminary data indicates that insertional mutagenesis of the vector at the RIAN locus, with subsequent dysregulation of local mRNA expression, occurs in AAV-associated HCCs. Determining why AAV gene delivery is infrequently associated with tumorigenesis in mice and whether such toxicity is relevant to humans remain as unresolved questions for those using AAV in human gene therapy applications.
2299T
Developing resources to alleviate muscle atrophy in FSHD by gene engineering. S. Das, B.P. Chadwick. Department of Biological Science, The Florida State University, Tallahassee, FL.

Facioscapulohumeral muscular dystrophy (FSHD), a debilitating disease with currently no cure or effective therapy, is the third most common inherited form of muscular dystrophy. FSHD is primarily characterized by progressive weakness and atrophy of skeletal muscle of the face, shoulders and upper arms. Both forms of this autosomal dominant disorder (FSHD1 and FSHD2) are likely caused by the reactivation of the double homeobox gene DUX4 that is epigenetically repressed in somatic cells. The more common form, FSHD1, is linked to a specific number of repeat units of the macrosatellite D4Z4 at 4q35. Each repeat unit contains the DUX4 open reading frame. FSHD2 is a result of a digenic inheritance of mutant SMCHD1 gene and a permissive 4q allele but is contraction-independent. The inappropriate expression of DUX4 in somatic muscle inhibits muscle differentiation and replenishment. Importantly, both forms of FSHD have a strict requirement for specific allelic variants of chromosome 4q and at least one unit of D4Z4, as complete loss of the repeat does not result in disease. Taking advantage of this, we are focused on development of an effective therapeutic strategy for FSHD using genome engineering in human cells through custom-built Transcription Activator Like Effector Nucleases (TALENs). Currently, we are developing TALEN targets to treat regions immediately proximal and distal to the D4Z4 array that can be used to remove the array or disrupt the polyadenylation signal on a permissive chromosome, in order to alleviate toxic gain of function. We are currently testing candidate nucleases on model cell lines. Successful editing of the diseased allele will permit use of our custom-built nucleases, to modify patient specific induced pluripotent stem cells (iPSCs) to achieve a therapeutic goal of using the same cells for the therapeutic strategy for muscle restoration. Our approach could be used in FSHD patient-specific cell transplantation therapy, in an attempt to alleviate this debilitating condition, and improve quality of life.

2300F

There was a reduction in GM3 ganglioside in the IT-treated dogs by immunochemistry. There was a 2.5-fold normal for AAV9 and AAVrh10, respectively. GUSB activity in the cisterna magna with the other vector (4E12GC/kg) of either an AAV9 (n=2) or AAVrh10 (n=2) vector encoding the MPS VII dogs was treated IV at post-natal day 3 with 2E13 genome copies (GC)/kg of either an AAV9 (n=2) or AAVrh10 (n=2) vector encoding the MPS VII dogs. Enzyme activity in the CNS from the IV-only injected dogs was <4% of normal. Only after IT delivery were low levels of AAV-antibodies detected in the CSF. Histochemically determined that IT injection of AAV vectors resulted in impressive levels of expression in the CNS of MPS VII dogs and has tremendous potential for treating the CNS manifestations of lysosomal storage disorders.

2301T

Spinal muscular atrophy (SMA) is a pediatric neuromuscular disease caused by mutations in SMN1. Previously, we showed that delivery of an AAV vector encoding SMN1 into the CNS of a mouse model of SMA corrected several aspects of the disease and afforded a significant extension in their lifespan. Here, we determined (i) the minimal number of genetically modified motor neurons in the spinal cord of SMA mice necessary for clinically relevant efficacy, and (ii) the level of motor neuron transduction efficiency following intrathecal delivery in juvenile pigs and non-human primates. Injection of 5e10, 1e10, and 1e9 genome copies (gc) of scAAV9-SMN1 into the CNS of SMA mice extended their median lifespans from 14 days to 153d, 70d, and 18d, respectively. Analysis of the spinal cords treated with 5e10, 1e10, or 1e9 gc showed that 30-60%, 10-30%, and <5% of the motor neurons had detectable levels of hSMN protein, respectively. Hence, achieving at least 10-30% motor neuron transduction was sufficient to significantly improve motor function and survival in SMA mice. To determine the results of gene transfer in larger animal models, a viral vector encoding GFP (scAAV9-GFP) was administrated intrathecally into juvenile pigs and monkeys. Analysis showed that a motor neuron transduction rate as high as 25-75% throughout the spinal cord in these large animal studies. This minimum number of genetically modified motor neurons appears to be necessary for clinical efficacy in SMA mice could be obtained in large animal models, justifying the continual development of intrathecal gene therapy for SMA.

2302F
Treatment of MFRP (Membrane frizzled-related protein)-related degeneration in patient-specific stem cells and a preclinical mouse model. Y. Li, Y.T. Tsai, C.W. Hsu, W.H. Wu, S.H. Tsang. Edward Harkness Eye Institute, Columbia University, New York, NY.

In the current era of personalized medicine, a large number of genetic variants have been discovered in patients with various diseases using next-generation sequencing techniques. Traditionally, to prove that genetic variants cause diseases, we have had to rely on animal models. However, substantial differences exist between mice and humans, including but not limited to drastic differences in lifespans. For instance, dopaminergic neuron projections are shorter in relation to overall body length in mice than in humans; thus, alpha synuclein SncA knockout mice do not develop Parkinson’s disease. Instead, as new reprogramming technologies have developed, it has become more feasible to generate patient-specific stem cell lines to validate sequence variants, elucidate pathophysiology, and perform targeted drug screening. We have used such reprogramming technologies to prove that a novel mutation in the MFRP (membrane-type frizzled-related protein) gene is pathogenic. Light microscopic images of human stem-cell-derived retinal cells taken from a patient displayed dysmorphic cells. Correction of the defect restored the patient’s retinal cells’ morphology and function. Similarly, correcting genetic defects in the Mfrp/+ or Mfrp–/– mouse can restore visual function in this preclinical model of retinitis pigmentosa.
Gene transfer and gene expression are attractive methods for the generation and the delivery of recombinant therapeutic proteins, which basically need an efficient and safe delivery tool. Viruses have been used for many years as gene transfer vehicle for the delivery and the transient expression of recombinant proteins. However, to date, there is no efficient non-viral delivery tool for expression of plasmids (pDNA) as compared to viruses. It is proposed that the unsuccessful development of non-viral vector may be due to many barriers such as the transfection efficiency and plasmid stability. It has been reported that viruses contain cis-acting RNA elements that facilitate the posttranscriptional processing and the exporting of mRNA. Furthermore, it was also reported that the DNA cis-acting matrix attachment region elements (MAR) promote homologous recombination to enhance the level of DNA stability in the cell. Moreover, BGH poly signal and/or matrix attachment region (MAR) in a non-viral vector may prolong the stability and enhance the expression of the delivered gene. Therefore, in this study, we constructed various plasmid vectors harboring different cis-acting elements that were evaluated for enhancing plasmid DNA as well as mRNA stability in two mammalian cell-lines, HEK293 and HeLa. Both cell lines were transfected with equal copy number and plasmid DNA and mRNA copy numbers per cell were measured during three weeks. Data showed that the combination of multiple cis-acting elements has significant, unexpected effect on both plasmid DNA and mRNA stability. In summary, these results provide valuable information to improve non-viral vectors stability and transient gene expression in mammalian cells; with implications in both, gene expression studies, as well as gene therapy.

Expression of human GNE through adeno-associated virus mediated therapy delays progression of myopathy in the GNE myopathy mouse model. M.C. Malicdan1,2, T. Okada3, S. Takeda3, F. Funato1, M. Huizing1, I. Nomaka1, Y.K. Hayashi1, Z. Argov1, I. Nishino4, S. Mitrai-Rosenbaum3, S. Noguchi1, 1) MGB, NHGRI, National Institutes of Health, Bethesda, MD; 2) Department of Neuromuscular Research, National Institute of Neuroscience, NCNP, Tokyo, Japan; 3) Department of Molecular Therapy, National Institute of Neuroscience, NCNP, Tokyo, Japan; 4) Hadassah-Hebrew University Medical School, Jerusalem, Israel.

GNE myopathy, also known as distal myopathy with rimmed vacuoles (DMRV) or (hereditary inclusion body myopathy (hIBM) is an early adult, moderately progressive myopathy due to mutations in GNE. This gene encodes a bifunctional enzyme critical to synthesis of sialic acid. GNE myopathy so far has been shown to be a disorder due to reduced sialylation of certain glycoconjugates in tissues including the muscle, as replenishment phenotype. The caveat in sialic acid supplement therapy lies in the pharmacokinetic properties of sialic acid; its relatively short half-life due to rapid catabolism. Certain glycoconjugates in tissues including the muscle, as replenishment of endogenous nuclease, or non-degradative mechanisms, where ASO binding sterically blocks or modulate translation, capping, or splicing. Genetic disorders in which a dominant mutation results in a toxic gain of function, such as liver diseases associated with aggregates in liver and skeletal muscle are attractive to approach with ASO therapy. An ASO directed to TTR mRNA resulted in dramatic reductions in mutated TTR protein in mouse models and in normal volunteers. In addition, ASO targeting TTR is under evaluation as therapy for familial disease. ASO designed to target T1R1 mRNA results in significant reduction in CTR1 levels resulting in reduction in copper levels, in the blood, brain and liver and improved liver pathology in a mouse model of Wilson disease. Furthermore, antisense oligonucleotides can be effective therapies using non-degradative mechanisms. For example, ASO for hearing loss, designed to correct SMN2 splicing, restores SMN expression and extended life span in SMN mouse model. Many additional genetic diseases are successfully being targeted with ASO technology for example; ASO inhibition of apolipoprotein B synthesis by Kynamro (mipomersen sodium) is an effective therapy to reduce LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia. This presentation will cover antisense technology platform and will include data from preclinical research and clinical trials with ASO drugs treating multiple genomic disorders.
2307T

Huntington's disease (HD) is a fatal autosomal dominant neurodegenerative disease that results from an expansion of polyglutamine residues in the huntingtin (Htt) protein. With the identification of the underlying basis of HD, therapies are being developed that reduce the expression of the causative mutant Htt. RNA interference (RNAi) seeks to selectively reduce the expression of such disease-causing agents as emerging as a potential therapeutic strategy for treating dominant disorders. This study examines the merits of administering a recombinant adenovirus associated viral (AAV) vector designed to deliver a shRNA to target and lower Htt levels in the striatum, at correcting the biochemical and behavioral deficits associated with the YAC128 mouse model of HD. We showed that AAV-mediated shRNA expression partially reduced the levels of wild-type and mutant Htt in the striatum. Concomitant with these reductions were significant improvements in both motor and behavioral deficits in YAC128 mice. Lowering mutant Htt levels was also associated with a reduction in the number of Htt aggregates that accumulated in the striatum of YAC128 mice. These results are supportive of the development of AAV-mediated RNAi as a therapeutic strategy for the treatment of HD.

2308F

Our aim is to develop a novel technological platform to cure HIV disease. The major innovation is to use hematopoietic stem cell (HSC) transplantation of the cells resistant to HIV (such as CCR5 delta 32 cells) spontaneously occurring in 4-15% of EU population) while improving the outcome of engraftment of HIV resistant cells in HIV patients with AIDS lymphoma or leukemia by suppressing graft-versus-host disease using gp120. This is based on novel evidence for the cure of HIV infection by CCR5 delta 32 stem cell transplantation (‘Berlin patient’ reported in Alers et al. 2011, Blood 117: 2791) and by novel insights of the protection from graft-versus-host disease by HIV-1 envelope protein gp120-mediated in human regulatory T cells by elevated levels of cAMP (Becker et al. 2009, Blood 114: 1263; Vaeth et al. 2011, PNAS 108: 2480; Klein et al. 2012, J Immunol 186: 1091).

The concept of allogeneic stem cell transplantation of the cells resistant to HIV using regimen improving function of regulatory CD4+ T cells favours suppressing graft-versus-host disease while retaining the benefits of graft-versus-leukemia effect. Therefore, the concept constitutes an outstanding example of cure rather than indefinite anti-retroviral treatment. The unprecedented power of the hematopoietic stem cell transplantation of the CCR5 delta 32 cells resistant to HIV has been proven to cure HIV infection in the case reported last year (‘Berlin patient’). Successful reconstitution of CD4+ T cells at the systemic level as well as in the gut mucosal immune system was found while the patient remained without any sign of HIV infection. Furthermore, during the process of immune reconstitution, evidence for the replacement of long-lived host tissue cells with donor-derived cells indicates that the size of the viral reservoir has been reduced over the time. Next step is to create database of CCR5 delta 32 donors in Stem Cell Bank and further optimize treatment of graft-versus-host disease by gp120 in phase I clinical trials at Transplantation Unit. Clients are HIV patients with AIDS-related lymphoma or specific forms of leukemia. Side effects of HSC against HIV will be minor and restricted to graft-versus-host disease. HSC against HIV treatment costs are in the same range as the costs of HSC therapy, but considerably less expensive than lifelong anti-retroviral (ART) therapy.

2309T
Lipid storage and impaired function in iPSC and monocyte - derived Gaucher macrophages are reversed with a non-inhibitory chaperone. E. Sidransky1, E. Alfaki1, N. Tayeb1, B. Stubblefield2, E. Maniwa2, G. Lopez2, E. Goldin2, S. Patnaik2, J. Marugan2.

1) Molec Neurogen, 1A213, MGB/NIH/NIH, Bethesda, MD; 2) NIH Chemical Genomics Center, NIH Center for Translational Therapeutics, National Institutes of Health, Rockville, MD, USA.

Gaucher disease (GD), the inherited deficiency of glucocerebrosidase, manifests with lysosomal glycolipid storage, primarily in macrophages. Chemical chaperone therapy has been developed as a means to increase glucocerebrosidase levels in patients. Since improved cell models of GD exhibiting a storage phenotype were needed to facilitate drug development, monocyte and induced pluripotent stem cell- (iPSC) derived-macrophages from patients with GD were generated and characterized. Both patient (n=20) and iPSC-derived macrophages (n=5 lines) share the same phenotype and exhibit reduced glucocerebrosidase activity. Storage, enhanced with tagged glycolipids and erythrocyte ghosts from patients with GD, was confirmed by lipid quantification and fluorescence microscopy. The Gaucher macrophages showed efficient phagocytosis, while effrocytosis of dead cells was impaired. Intracellular reactive oxygen species production was reduced and phagosomes maturation altered, indicating that Gaucher cells have defective digestion of phagocytes. Using high-throughput screening with patient spleen as the source of mutant enzyme, several non-inhibitory, small molecule chaperones for glucocerebrosidase were identified. The lead compound reversed the disease phenotype, enhanced glucocerebrosidase specific activity, reduced lipid storage and normalized the effrocytic index.

Our new capability to evaluate alterations in models closely mimicking the human disease should enable us to understand the disease pathogenesis and facilitate the development and characterization of new treatments for Gaucher disease.

2310F
MicroRNA-486 overexpression delays the disease pathology of dystrophin-deficient muscle. M.S. Alexander1, J.C. Casar2, N. Motohashi1, N.M. Vieira1, M.J. Gasperini1, J.A. Myers1, E.A. Estrella1, P.B. Kang1,2, F. Shapiro6, G. Kawahara1, E. Gussoni1, L.M. Kunkel1,2,3,4, 1) Genetics, Boston Children's Hospital/Harvard Medical School, Boston, MA; 2) The Manton Center for Orphan Disease Research at Boston Children's Hospital, Boston, MA; 3) Department of Pediatrics and Genetics at Harvard Medical School; 4) Harvard Stem Cell Institute, Cambridge, MA 02138; 5) Department of Neurology, Boston Children's Hospital and Harvard Medical School; 6) Departments of Orthopedic Surgery at Boston Children's Hospital and Harvard Medical School, Boston MA 02115.

Duchenne Muscular Dystrophy (DMD) is caused by mutations in the dystrophin gene that result in the dysregulation of many signaling pathways that interact directly or indirectly with the dystrophin protein. Previously, we identified miR-486 as being strongly reduced in its expression levels in the dystrophin-deficient mouse and muscle biopsies of human DMD patients. Here we report that transgenic overexpression of the muscle-enriched microRNA, miR-486, in mdx5cv (dystrophin-mutant) mice resulted in improved serum biochemistry, reduced apoptosis, increased myofiber size, and improved muscle regeneration following injury. Transient overexpression of miR-486 also resulted in similar improved muscle physiology in the mdx5cv mouse. Using a bioinformatic approach, we identified DOCK3, a dedicator-of-cytokinesis-3, as being a direct downstream target of miR-486 in skeletal muscle. Modulation of DOCK3 expression in myoblast cell culture had strong effects on normal and DMD myoblast proliferation, apoptosis, and differentiation capabilities. Together, these studies demonstrate that stable overexpression of miR-486 delays many of the signs of the disease pathology of dystrophin-deficient muscle.
2311T
Epigenetic upregulation of Survival of Motor Neuron 2 (SMN2) protein by oligonucleotides targeting long non-coding RNA. K.M. Hussey1, R. Davey1, E. Brand1, D. Knowlton1, B. Schwartz2, R. Subramanian1, J. McSwiggen1, A. Krieg1, J. Keil2, H. Cardona1, C. DiDonato1, J. Barsoum1.
1) RaNA Therapeutics, Cambridge, MA; 2) Northwestern University & Lurie Children’s Hospital of Chicago Research Center, Chicago, IL.

Spinal Muscular Atrophy (SMA) is a motor neuron disease affecting 1 in 6,000 children. This inherited disease leads to muscle weakness and atrophy and is mainly caused by a deleted or defective survival of motor neuron 1 (SMN1) gene. The SMN1 (telomeric) protein is a key player in cellular splicing events as a member of the snRNP complex in spinal cord neurons. A nearly identical gene, SMN2 (centromeric), is expressed in the same cells, albeit at very low levels due to a single point mutation that results in defective splicing. SMN2 is able to functionally compensate for SMN1 and support a milder phenotype when SMN2 copy number is increased, suggesting SMN2 as a primary therapeutic target. RaNA’s goal is to specifically up-regulate the transcripts derived from the SMN2 gene, ultimately increasing functional SMN2 protein and improving SMA disease conditions.

Our platform is based on a recently discovered class of long noncoding RNAs (lncRNAs) that interact with the Polycomb Repressive Complex 2 (PRC2) and act in cis to repress gene transcription. RaNA’s approach is to utilize short synthetic oligonucleotides (oligos) to interfere with this lncRNA-PRC2 interaction specifically at the SMN2 locus, relieve SMN2 repression, and drive transcription of the SMN2 gene. To this end, we have identified several oligos that up-regulate full length SMN2 mRNA 3-5 fold and protein by 3-7 fold above baseline in SMA patient-derived fibroblasts. These oligos target lncRNA transcripts in the SMN2 gene locus. Our results indicate that RaNA oligos enhance transcription of full length SMN2 mRNA and functional protein as well as correct the defective splicing of the transcript. Preliminary studies in a humanized SMA mouse model suggest that RaNA oligos can positively affect preclinical outcomes.

Current in vivo studies are focused on determining the effects of RaNA oligos on various preclinical phenotypes and survival in humanized SMA mouse models. Furthermore, we are investigating global- and locus-specific epigenetic and transcriptional changes to understand the mechanisms of RaNA oligo action. Collectively, our data suggests a novel epigenetic mechanism to increase SMN2 protein production that will likely provide a significant therapeutic benefit to SMA patients.

2312F
Towards the therapy for Angelman syndrome. L. Meng1, A. Award2, F. Rigo3, A. Beaudet1.
1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Isis Pharmaceuticals, Carlsbad, CA.

Angelman syndrome (AS) is a severe neurodevelopmental disorder caused by maternal deficiency of the imprinted gene UBE3A. Although the molecular mechanisms of maternal deficiency are diverse, all AS patients carry at least one copy of paternal UBE3A, which is silenced but intact. By correcting the expression level of UBE3A via activating the silenced paternal allele, the disease might be treated. UBE3A-ATS is the antisense transcript of UBE3A that negatively regulates its expression. By truncating this antisense RNA in mice, we showed that depletion of Ube3a-ATS is sufficient to activate expression of paternal Ube3a and rescue phenotypic defects in the Angelman syndrome mouse model, including motor defects, cognitive deficit, and impaired long-term potentiation. Next, to develop specific therapy for Angelman syndrome, we sought to activate paternal Ube3a by knocking down Ube3a-ATS with antisense oligonucleotides (ASOs). With screens performed in cultured mouse neurons, we successfully identified several mouse specific ASOs that can achieve full activation of paternal Ube3a. Different from topoisomerase inhibitors, ASOs we identified are more site-specific, with little or minimal effect on nearby paternally expressed genes of Snord116 and Snord117. Compared with PBS-treated animals, mice received intracerebroventricular injection of ASOs show significant knock-down of Ube3a-ATS and activation of paternal Ube3a in the cortex and other brain regions four weeks post injection, indicating that the ASOs function effectively in vivo. Future studies aim to achieve phenotypic improvement in the Angelman syndrome mouse model following CNS administration of ASOs and identifying human-specific ASOs for treating Angelman syndrome.
2313T Decoding the biochemistry of Snyder-Robinson Syndrome: Using yeast to model the polyamine biosynthetic defect. J.S. Albert1, M.K. Chattopadhyay2, W. Bone1, L.A. Wolfe1, W.A. Gahl1, C.F. Boerkoel1, 1) NHGRI, National Institutes of Health, Bethesda, MD; 2) NIDDK, National Institutes of Health, Bethesda, MD.

Polyamines are simple, positively charged, ubiquitous molecules that interact with anionic compounds such as DNA, RNA, and ATP. Homeostasis of the polyamines putrescine, spermidine, and spermine is essential to cell growth and survival and is maintained primarily by the polyamine biosynthetic enzymes ornithine decarboxylase (ODC), S-adenosylmethionine decarboxylase (AdoMetDC), spermidine synthase (SRM) and spermine synthase (SMS). The balance of spermine and spermidine is crucial for proper chromatin structure, ion channel regulation, transcription and translation. Disruptions in polyamine homeostasis are associated with Snyder-Robinson syndrome (SRS), an X-linked intellectual disability syndrome first reported in 1969. SRS results from a mutation in SMS1 and produces a complex multi-systemic phenotype including intellectual disability, dysmorphic facies, muscle hypotonia, kyphoscoliosis, osteoporosis, seizures, and speech and gait abnormalities. The mechanism whereby dysregulation of polyamine homeostasis results in SRS is unknown. We are using an S. cerevisiae knockout model of SMS (spe4Δ) to study the polyamine biosynthetic pathway. Using HPLC analysis, we have found that SPE4 knockout yeast strains, compared to wild-type, have only a 1.3 fold increase in spermidine under physiologic conditions and only a 1.5 fold increase in spermidine in the presence of superphysiologic (10mM) spermidine concentrations. This attenuated increase in spermidine is similar to that observed in individuals with SRS. Spermine is undetectable in the SPE4 knockout strain. We also found elevated levels of methyldihydrospermidine (MTPA), a product of the biosynthetic pathway and an inhibitor of the upstream enzyme, spermidine synthase, in the cell extract of the SPE4 knockout, but not in the media. This suggests that increase in MTPA could serve to protect the cells from accumulating deleterious amounts of spermidine. These results suggest that accumulation of spermidine and deficiency of spermine are not the cause of SRS. We discuss this as well as our ongoing studies on the possible biochemical etiologies for SRS.

2314F Secondary Coenzyme Q10 deficiency in patients with BRAF mutations. E. Trevisson1, M.A. Desbats1, C. Cerqua2, M. Doimo1, A. Casarin1, C. Santos-Ocana2, P. Navas Lloret2, L. Salviati1. 1) Woman and Child Health, Clinica Genetica Unit, Univ Padova, Padova, Italy; 2) Centro Andaluz de Biologia del Desarrollo, Universidad Pablo de Olavide, Sevilla, Spain.

Coenzyme Q10 (CoQ) is a small lipophilic molecule that plays fundamental cellular functions: it acts as an electron transporter between complexes I and II of the mitochondrial respiratory chain and complex III, it functions as a hydrogen ion channel in the myoneural junction, it acts with anionic compounds such as DNA, RNA, and ATP, and it has a role in ion channel regulation, transcription and translation. Disruptions in polyamine homeostasis are associated with Snyder-Robinson syndrome (SRS), an X-linked intellectual disability syndrome first reported in 1969. SRS results from a mutation in SMS1 and produces a complex multi-systemic phenotype including intellectual disability, dysmorphic facies, muscle hypotonia, kyphoscoliosis, osteoporosis, seizures, and speech and gait abnormalities. The mechanism whereby dysregulation of polyamine homeostasis results in SRS is unknown. We are using an S. cerevisiae knockout model of SMS (spe4Δ) to study the polyamine biosynthetic pathway. Using HPLC analysis, we have found that SPE4 knockout yeast strains, compared to wild-type, have only a 1.3 fold increase in spermidine under physiologic conditions and only a 1.5 fold increase in spermidine in the presence of superphysiologic (10mM) spermidine concentrations. This attenuated increase in spermidine is similar to that observed in individuals with SRS. Spermine is undetectable in the SPE4 knockout strain. We also found elevated levels of methyldihydrospermidine (MTPA), a product of the biosynthetic pathway and an inhibitor of the upstream enzyme, spermidine synthase, in the cell extract of the SPE4 knockout, but not in the media. This suggests that increase in MTPA could serve to protect the cells from accumulating deleterious amounts of spermidine. These results suggest that accumulation of spermidine and deficiency of spermine are not the cause of SRS. We discuss this as well as our ongoing studies on the possible biochemical etiologies for SRS.

2315T Fine-mapping of the Atp7a promoter deletion and biochemical characterisation in the mottled-dappled mouse model of Menkes disease. M.R. Hadfield1, K. Patel1, P. Sullivan2, D.S. Goldstein3, J.A. Centeno4, S.G. Kaler1. 1) Molecular Medicine Program; NHCHD, NIH, Bethesda, MD; 2) Clinical Neurosciences Program, NINDS, NIH, Bethesda, MD; 3) Division of Biophysical Toxicology, Joint Center for Pathology, Malcolm Grow Medical Clinic, Andrews Air Force Base, Camp Springs, MD.

The mottled-dappled (mo-dp) is a mouse model of X-linked recessive Menkes disease caused by a large, previously uncharacterized deletion in the 5’ region of Atp7a, the mouse ortholog of human ATP7A. Affected males develop hydrocephalus, die between 3 and 5 days of age, and show bone and muscle abnormalities. The mottled-dappled mouse exhibits reduced expression of Atp7a in the lysosomes, and distorion of the pericellular and pelvic girdles and limb bones (Kaler, Nat Rev Neurol 2011). We have embarked on a preclinical approach (Haddad et al, Mol Ther Nucleic Acids 2013) for rescue of mo-dp to provide additional preclinical data in support of translation to a gene therapy approach for Menkes disease. To develop a reliable genotyping assay for the pre-clinical gene therapy study, we designed a custom 4x180K microarray on the mouse X chromosome and tiled 60-mer probes around the region of interest near the Atp7a locus. The average resolution (probe spacing) of the microarray in the vicinity of the deletion was 675 bp. We performed comparative genomic hybridization (CGH) using extracted DNA from normal and carrier mo-dp females. CGH revealed a deletion of ~9.4 kb. We fine-mapped the deletion breakpoints using PCR and successfully amplified a junction fragment of 630 bp using the following primers: F-TGTCAAGGTCTTCTGGGAGAT and R-CAGACACTGGGAGGATTGAAG. Sequencing the junction fragment disclosed the exact breakpoint positions and determined that the mo-dp deletion is precisely 8.245 kb and includes 2 kb in the promoter region, exon 1 and 2 and intron 1. Whole exome sequencing of mouse lymphoma cells from brains showed decreased amounts of full length Atp7a protein compared to normal, consistent with reduced expression due to the promoter region deletion on one allele. In addition, brain neurochemicals from heterozygous female mice showed statistically significant higher DA/NE and DOPAC/DHPG ratios compared to normal (p=0.0128 and 0.0034, respectively), consistent with partial deficiency of dopamine-beta-hydroxylase, a dopa-perdependent brain enzyme that converts dopamine to norepinephrine. Brain copper measurements are pending. Our mouse model will improve the correlation of genotype with biochemical phenotype, as well as understanding of treatment response in the context of a null Atp7a mutation.


Gaucher Disease (GD), the most prevalent lysosomal storage disorder, is caused by mutations in GBA1, the gene encoding glucocerebrosidase (GCase). The disease is characterized by the presence of glucosylceramide (GlcCer)-laden macrophages in the liver, spleen and bone marrow. Although in GD GCase is deficient in all cell types, the phenotype manifests primarily in macrophages. There is a dynamic relationship between phagocytosis and autophagy, the evolutionarily ancient process of lysosomal self-digestion of organelles, apoptotic corpses and cytosolic pathogens. We investigated autophagy in macrophages derived from peripheral monocytes from ten patients with type 1 GD (G-Macs) and two induced pluripotent stem cell (IPSC)-derived Gaucher macrophage lines, all sharing the common GD genotype N370S/N370S. Increased P62 expression in these macrophages parallels increased autophagy. Levels of both LC3-II and P62 were increased in G-Macs during etrophagy and inflammation, as demonstrated by inflammasome activation with LPS and ATP. Moreover, expression of VPS34 and Rab7, proteins important in the fusion of early endosomes to late endosomes, decreased significantly. Autophagosomes deliver phagosomes to lysosomes for degradation, whereas inflammasomes are activated by infection, which in turn regulates the activity of caspase-1 and the maturation of interleukin (IL)-1 beta and IL-18. Macrophage polarization, evaluated using PCR arrays, indicates that G-Macs manifest the alternatively activated phenotype in the presence of LPS with ATP. Importantly, G-Macs show increased expression of IL-18 and IL-10, but decreased IL-1 beta expression. Our data indicates that impaired secretion of some inflammatory cytokines results from the suppression of autophagy, resulting in increased Igf16L in GD, which leads to an increased histone acetylation of the IL-1β promoter in these cells in the presence of LPS and ATP. This study demonstrates that lysosomal dysfunction leads to the suppression of autophagy, followed by inhibition of inflammasome activity.
2317T  Resistin increases the expression of aggrecanases in human chondrocytes. K.O. Yaykasli1, E. Yaykasli2, E. Kayar3, M. Ozsahin4, M. Uslu1, 1) Department of Medical Genetics, Medical Faculty, Duzce University, Duzce, Turkey; 2) Department of Medical Biology and Genetics, Institute of Health Science, Duzce University, Duzce, Turkey; 3) Department of Medical Pharmacology, Medical Faculty, Duzce University, Duzce, Turkey; 4) Department of Physical Medicine and Rehabilitation, Medical Faculty, Duzce University, Duzce, Turkey; 5) Department of Orthopedics and Traumatology, Medical Faculty, Duzce University, Duzce, Turkey.

Objective: The adipokine, resistin, was discovered in 2001 and expressed by not only adipose tissue but also by macrophages in high levels. First, resistin has been linked to insulin resistance, but further studies have linked it to other physiological processes, such as inflammation and immunity. Rheumatoid arthritis (RA) is an inflammatory and autoimmune disease that is characterized by irreversible destruction of the extracellular matrix (ECM). Aggrecanases are the primary enzymes responsible for the degradation of ECM. Several current investigations have found that elevated resistin levels are associated with aggrecanase expression. The aim of this study was to investigate the putative roles of resistin in aggrecanases upregulation. Material and Method: Human articular chondrocytes were cultured with resistin at ten concentrations (100, 250, and 500 ng/ml). After incubation for 12h and 24h, total RNA was extracted using TriPure reagent, and 2 μg of total RNA was reverse-transcribed by random primer. The ADAMTS1, ADAMTS4, ADAMTS5, ADAMTS9, and β-actin genes expression were analyzed by quantitative real-time polymerase chain reaction. Results and Conclusion: Resistin increased the expression of ADAMTS1, ADAMTS4, ADAMTS5, and ADAMTS9 genes at 100 ng/ml and 250 ng/ml doses. The irreversible damage of ECM by aggrecanases is a key process in RA. An elucidation of the underlying causes for elevated levels of aggrecanases may be beneficial to understanding the etiology of RA. In conclusion, resistin with inflammatory properties is involved in RA by increasing the expression level of aggrecanases. However, further studies are needed to clarify the action mechanism of resistin.


Patients with type I Gaucher disease (GD1) present not only with hematological and visceral manifestations, but multiple bone abnormalities including osteopenia, Erlenmeyer flask deformity, bone crisis, and avascular osteonecrosis. A recent mouse model of GD1 generated via conditional knockout of glucocerebrosidase (Gba1) in hematopoietic and mesenchymal cells displayed GD1-like bone pathophysiology including osteopenia, reductions in bone formation rate, and osteonecrosis. The skeletal phenotype present in this transgenic model was attributed to defects in both osteoblast proliferation and differentiation. Given the potential relationship between glucosylceramide (GL-1) and osteoblast activity, we sought to determine the impact of inhibiting GL-1 synthesis upon bone physiology in mice. We therefore administered the glucosylceramide synthase (GCS) inhibitor Genz-667161 (60 mg/kg/day) in chow to C57Bl/6 mice for a period of 6 weeks and subsequently examined trabecular bone in femurs and lumbar vertebrae via histomorphometric analysis. In distal femur we observed an increase in bone volume fraction with a concomitant decrease in trabecular spacing in mice treated with Genz-667161. In L4 lumbar vertebrae we also noted a decrease in trabecular spacing with an associated increase in trabecular number. These effects in lumbar vertebrae would be expected to provide an increase in bone strength. This preliminary data suggests that GL-1 and perhaps other glycosphingolipids may play a role in beneficial bone remodeling and provide the rationale for further studies that examine the effect of GL-1 upon bone formation and resorption.

2319T  Fat mass and obesity-associated (FTO) Protein interacts with CamKII and modulates the activity of CREB signaling pathway. L. Lin, P. Jin.

Human Genetics, Emory University School of Medicine, Atlanta, GA.

Polyphenomics in the fat mass and obesity-associated (FTO) gene have been associated with obesity in humans. Alterations in FTO expression in transgenic animals affect body weight, energy expenditure and food intake. FTO is a nuclear protein and its physiological function of FTO remains largely unknown. To understand the molecular functions of FTO, we performed yeast two-hybrid screen to identify the protein(s) that could directly interact with human FTO protein. Using multiple assays, we demonstrated that FTO interacts with CamKII. FTO interacted with three isoforms of CamKII: α, β, and γ. A recent mouse model of GD1 generated via conditional knockout of glucocerebrosidase (Gba1) in hematopoietic and mesenchymal cells displayed GD1-like bone pathophysiology including osteopenia, Erlenmeyer flask deformity, bone crisis, and avascular osteonecrosis. We found that the overexpression of FTO could delay the dephosphorylation of CREB in human neuroblastoma (SK-N-SH) cells.

Expression levels of NPY1 and BDNF, which are targets of CREB, were dramatically increased due to the prolongation of CREB phosphorylation in the presence of FTO. Both NPY1 and BDNF have been shown to regulate food intake and energy homeostasis. Thus, our results suggested that FTO could potentially modulate obesity by regulating the activity of CREB signaling pathway.


1) Department of Pediatrics and Child Health, Kurume University, Kurume, Japan; 2) Research Institute of Medical Mass Spectrometry, Kurume University, Kurume, Japan; 3) Department of Human Genetics, Nagasaki University, Nagasaki, Japan; 4) Genetics Division, Department of Pediatrics, LAC+USC Medical Center, University of Southern California, Los Angeles, CA, USA.

Background: Inborn errors of energy metabolism including amino acid and fatty acid metabolism defects have recently become disorders screened by tandem mass spectrometry based newborn screening. The majority of patients with these disorders show positive results with characteristic profiles suggestive of a specific diagnosis. On the contrary, patients with carbohydrate metabolism and mitochondrial respiratory chain complex (RCC) defects often do not show positive newborn screening results with characteristic profiles. Hepatomegaly, hypoglycemia, ketosis, and lactic acidosis can be seen in patients with carbohydrate metabolism and RCC defects. We had a patient with infantile onset hypotonia, chronic massive hepatoemegaly, hypercholesterolemia, chronic ketosis, mild developmental delay and episodic lactic acidosis: No diagnosis could be made on the enzyme assays.

Objective: To recruit and perform exome studies on patients with similar clinical findings. Case report: The patient was a 6-month-old Japanese male with hepatomegaly (fatty liver), lactic acidosis, hyperlipidemia, hypoglycemia, ketonuria, and mild developmental delay. The patient was born to a 26-year-old Japanese mother and a consanguineous Japanese father (half sibling through their mother). Pregnancy was reportedly unremarkable. The patient was delivered at 37 weeks and weighed 2940g. Apgar scores were 8 and 9. The patient was fed with breast milk. At age 5 months, generalized mild hypotonia, failure to thrive and hepatomegaly (the edge was at 14 cm below RCM) were noted by his primary pediatrician. The following laboratory studies were performed: AST 267, ALT 57, Plasma amino acids-unremarkable. The patient had mild ketosis, plasma acylcarnitine- C2 (40.9 μM) and C6 (0.5 μM), Ca 12mg/dl, PO4 2.9mg/dl, FFA 1791μM, total ketones 2849μM, 3OH 2211μM/AA 638μM, blood lactate 118mg/dl, uric acid 1.14mg/dl. Enzyme assays for glycogen storage diseases including phos- phorylase, debranching enzyme, and phosphorylase b kinase were unremarkable. Liver biopsy at age 7 months showed severe fatty liver of unknown etiology. Mitochondrial respiratory chain assays with liver tissues were unremarkable. Discussion: The patient’s presentation was suggestive of glycolytic pathway defects including fructose metabolism defects. Exome studies may identify the gene(s) responsible for the findings.
**2321T**

Delineation of renal-independent disease mechanism underlying Hypophosphatemic Familial Tumoral Calcification caused by GALNT3 mutations. Y. Yang1,2, G. Stork1,2, A. Kassar1,2, R. Y. Kato1,2, A. P. Gaffney1,2, N. Bhattacharya1,2, R. L. Gaffney1,2, W. A. Gaith2,3, H. C. Dietz1,4,5,1. 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Genetic Disease Research Branch, National Human Genome Research Institute, Bethesda, MD; 3) Craniofacial and Skeletal Diseases Branch, National Institute of Dental and Craniofacial Research, Bethesda, MD; 4) Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD; 5) Undiagnosed Diseases Program, National Institutes of Health, Bethesda, MD; 6) Howard Hughes Medical Institute, Chevy Chase, MD.

Hyperphosphatemic Familial Tumoral Calcification (HFTC) is characterized by progressive deposition of calcium phosphate complexes in soft connective tissues and periarticular tissues; serum phosphate levels are markedly increased. Loss-of-function mutations in three genes (FGF23, GALNT3, and KLOTHO) involved in renal phosphate excretion have been implicated in the pathogenesis of HFTC. FGF23, which is stabilized via O-glycosylation by GALNT3, is a circulating hormone that, with its co-factor klotho, decreases the expression of inorganic phosphate (Pi) transporter genes. Npt2A and Npt2C, in renal proximal tubule cells thus reducing Pi reabsorption. It has long been assumed that the calcification seen in HFTC is due to high levels of circulating Pi, which can then complex with calcium and precipitate. Here, we demonstrate that this calcification process can occur in an isolated, in vitro model system of ectopic calcification. Primary fibroblasts derived from two unrelated HFTC patients with biallelic GALNT3 mutations can calcify in vitro when stimulated with osteogenic media. These patients' cells also have increased tissue non-specific alkaline phosphatase (TNAP) and ENPP1 gene expression and enzyme activity. TNAP converts pyrophosphate (PPi) to inorganic phosphate (Pi), a negative regulator of calcification, while ENPP1 hydrolyzes ATP to AMP and PPi, a positive regulator of calcification. We have previously found that regulation of the Pi:PPi ratio may influence TNAP and ENPP1 levels, contributing to homeostasis. We found that HFTC cells have increased Gli-1 expression, a transcription factor that binds upstream of TNAP and increases target gene expression. To further explore the role of Gli-1 in ectopic calcification, we treated patient cells with GANT58, a Gli-1 inhibitor. Gli-1 inhibition led to increased cell calcification. Further delineation of this pathway will help us better understand the basic mechanisms of ectopic calcification and potentially reveal new treatment strategies for HFTC and perhaps related disorders.

**2322F**

The mitochondrial F-box protein FBXL4 is necessary for efficient mitochondrial respiration. C. A. Biagosto1,2, X. Gai3, D. Ghezzi4, M. A. Johnson5,6, H. E. Shamseddin5,6, T. B. Haack1,2, A. Reyes5,6, M. Tsukikawa2, C. A. Sheldon7, S. Srinivasan8,9, M. Gorza2,9, L. S. Kramer1,2, T. M. Strom1,2, E. Place1,2, S. Vidonis5,6, L. Wong10, M. Salih11, E. Al-Jishi12,13, C. P. Raab14, F. Furlan12,14,15, J. A. Mayr16, V. Konstantopoulou16, M. Huemer1,2, T. Melting1,2, P. Frei17,18, W. Späth18, F. S. Akkura16,17,18, M. J. Falk17, M. Zeviani1,2, P. Prokisch1,2. 1) Institute of Human Genetics, Technische Universität München, Munich, 81675 Germany; 2) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, 85764 Germany; 3) Department of Molecular Pharmacology and Therapeutics, Loyola University Stritch School of Medicine Maywood, IL 60153, USA; 4) Department of Molecular Neurogenetics, Institute of Neurology Besta, 23888 Milan, Italy; 5) MRC Mitochondrial Biology Unit, CB2 0XY, Cambridge, UK; 6) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 7) Division of Human Genetics and Metabolic Disease, Department of Pediatrics, Children's Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA; 8) Department of Animal Biology, University of Pennsylvania School of Veterinary Medicine, Philadelphia, PA 19104, USA; 9) Ocular Genomics Institute, Massachusetts Eye and Ear Infirmary (MEEI), Harvard Medical School, Boston, MA 02114, USA; 10) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA; 11) Department of Pediatrics, King Khalid University Hospital and College of Medicine, King Saud University, Riyadh, Saudi Arabia; 12) Salmaniyah Medical Complex, Arabian Gulf University, Bahrain; 13) Department of Pediatrics, Nemours/ Al DuPont Hospital for Children, Thomas Jefferson University, Wilmington, DE 19803, USA; 14) Unit of Human Genetics, Division of Pediatrics, Foundation MBBM/ San Gerardo University Hospital, 20900 Monza, Italy; 15) Department of Paediatrics, Paracelsus Medical University Salzburg, Salzburg 5020, Austria; 16) Department of Pediatrics, Medical University of Vienna, Vienna 1090, Austria; 17) Department of Pediatrics, Klinikum Reutlingen, Reutlingen, 72764 Germany; 18) Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.

Mitochondrial F-box protein FBXL4 encodes a member of the F-box family of proteins, which are involved in phosphorylation-dependent ubiquitination and/or G-protein receptor coupling, thought to be localized in the nucleus and cytosol. Surprisingly, we could localize FBXL4 to mitochondria using immunofluorescence and subcellular fractionation. Muscle and fibroblast tissue of patients displayed reduced activities of all respiratory chain complexes as well as mDNA depletion. Biochemical rescue experiments based on the lentiviral-mediated expression of wildtype FBXL4 was used to restore reduced oxygen consumption rate of OXPHOS-defective fibroblast cell lines. All individuals manifested early-onset lactic acidemia, hypotonia and developmental delay caused by severe encephalopathy, whereas some have presented with dysmorphism, skeletal abnormalities, poor growth, gastrointestinal dysmotility, renal tubular acidosis, seizures and episodic metabolic failure. So far, the results suggest a role of FBXL4 in organell maintenance, which is currently under investigation.

Mitochondria synthesize ATP through oxidative phosphorylation in all eukaryotic cells. Unlike other organelles, mitochondria contain their own genome, mtDNA, which in humans encodes 11 messenger-RNAs, 2 ribosomal-RNAs and 22 transfer-RNAs. The messenger RNAs are translated into 13 polypeptides, which are essential for the assembly of OXPHOS complexes (C I, III, IV and V). Transcription of mtDNA generates long, polycistronic transcripts which are processed by mitochondrial RNaseP (mt-RNaseP) and RNaseZ at the 5'- and 3'-ends of tRNAs, respectively. RNaseP is composed of three subunits: MRPP1, MRPP2 and MRPP3. Knockdown of either subunit results in inactivation of RNaseP and a decrease in the levels of mature tRNAs, mtRNAs and rRNAs. Interestingly, MRPP1 and MRPP2 also function to establish N1-methylation of purines at position 9 in several tRNAs. Notably, mutations in the methyltransferase domain of MRPP1, though eliminating its methylase activity, do not affect the processing activity of RNaseP indicating that both functions are independent.

We used whole-exome sequencing in order to identify mutations in novel nuclear genes associated with mitochondrial dysfunction. We focused on a series of patients with multiple respiratory chain deficiencies with no qualitative or quantitative mtDNA anomaly. This approach enabled us to identify a patient carrying a homozygous mutation in the MRPP1 gene. This patient is the only girl born to consanguineous parents of Turkish origin. She presented feeding difficulties at one month of age. She then progressively developed trunk hypotonia, nystagmus and hypertrophic cardiomyopathy. She had CI and CIV deficiency was examined from autopsy of a chronic neuroenopathic Gaucher disease patient. The nearly complete correction of the hepatic, splenic, and bone marrow disease contrasted with the large clusters of CD68 positive Gaucher cells in the interstitia/alveolae of lungs, and, ubiquitously, within the perivascular spaces of brain blood vessels, i.e., macrophages. In the brain, a few such macrophages were within the cerebellar, basal ganglia and hippocampus parenchyma. Only few perivascular macrophages were positive for the macrophage mannose receptor (MMR). In addition to large losses of neurons and absence of Purkinje cells, the above brain regions were positive for GFAP (astrogliosis), phosphorylated Tau, and synuclein, indicating proinflammation and neurodegeneration. GluCer and GluS accumulations were in massive mesenteric lymph nodes and lungs, but not in the liver. In lung and lymph nodes, major GluCer species were GluCer 24:0 and GluCer 16:0. In brain, basal ganglia, and hippocampus, GluCer species contained the long acyl chain GluCer 24:0 and GluCer 24:1, indicative of their visceral origin and presence in the predominant perivascular Gaucher cells that were unaffected by ET. These unique data show that long-acting ET with marnose-terminated GCase was effective for liver, spleen, and bone marrow. In comparison, ET had limited access to lung, lymph nodes, and brain. Surprisingly, the perivascular storage cells, which are on the vascular side of the blood brain barrier, were negative for the MMR, thereby blunting the effects of MMR-targeted ET, which has significant impact for therapies on either side of the blood brain barrier. In particular, the extra- and intra-parenchymal CNS Gaucher cells indicates the need for non-ET based therapies for global correction in the brain.
2326F
Characterization of an ACAD10 Deficient Mouse Model: Pathological and Biochemical Analyses. K. Kormanik1, D. El Demellawy2, A-W. Mohsen1, A. Karunanidhi3, M. Reyes-Mugica1, J. Vockley1, 2 1 University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA 15224; 2 University of Pittsburgh School of Medicine, Pittsburgh, PA 15224. 
Acyl-CoA dehydrogenase 10 transcript (ACAD10) codes for 1059-aa peptide with a ~400-aa C-terminus domain homologous to members of the ACAD flavoenzymes family involved in oxidation of fatty acids and intermediates of aa metabolism. A polymorphism in the ACAD10 gene has been linked to obesity in Pima Indians, but its physiological role remains unknown. The objective is to identify the role of ACAD10 in physiology and its link to diabetes. We have generated an ACAD10 deficient mouse and an anti-ACAD10 antibody to conduct pathological, biochemical, and molecular studies. Deficient animals are viable and fertile, but accumulate excess abdominal adipose tissue. Pathological studies reveal that ACAD10 deficient mice manifest an early inflammatory liver process and secondary splenic extramedullary hematopoiesis. Although skeletal muscle is histologically normal, deficient mice have elevated creatine kinase when fasted indicative of rhabdomyolysis. Metabolomics analysis identifies elevated levels of malonyl carnitine, succinyl carnitine, methylmalonyl carnitine, glutarylcarnitine, adipoylcarnitine in deficient mouse urine samples. Immunological studies reveals antigen detected in pancreas, lung and brain and localized to peroxisomes. RNAseq analysis reveals a broad spectrum of changes in gene expression including genes involved in the response to oxidative stress. We’ve made significant progress towards identifying the role of ACAD10 in physiology.

2327T

The mitochondrial trifunctional protein (MTP) is a hetero-octamer composed of the four α- and four β-subunits, which catalyzes the final three steps of mitochondrial β-oxidation of long chain fatty acids. To date, only two cases with MTP deficiency have been reported to be associated with hypoparathyroidism and peripheral polyneuropathy. We report that sibling cases born to consanguineous parents have MTP deficiency associated with infantile onset hypoparathyroidism and peripheral polyneuropathy. Genomewide linkage analysis and array CGH analysis eliminated known diseases of α-ketoglutarate dehydrogenase and branched-chain ketoacid dehydrogenase function. Five patients have been reported with thiamine pyrophosphokinase-1 (TPK1) deficiency, with encephalopathy, extrapyramidal, bulbar, and cerebellar signs, and metabolic decomposition leading to death. Some patients received high dose thiamine therapy. Patients: We present a sibling pair with Leigh’s disease, progressive hypotonia, regression, and chronic encephalopathy. Whole exome sequencing in the younger sibling demonstrated homozygous TPK1 mutations. The older sibling had died from progressive neurologic disease with metabolic strokes, and the younger sibling had progressive neurologic decline. MRI demonstrated putamen and thalamic abnormalities in the younger, with similar and progressive evolution in the older sibling. The patient was started on high dose thiamine, niacin, biotin, and α-lipoic acid; ketogenic diet was initiated to reduce metabolic burden. He subsequently demonstrated improvement in neurologic function, with reattainment of lost milestones. Conclusions: TPK1 deficiency should be considered in children with Leigh’s disease without identifiable mutations, and this gene should be evaluated as a candidate for testing in commercially available sequencing panels. Reducing metabolic burden in these children by addition of supplements and transition to ketogenic diet may reverse some of the neurologic progression and improve outcome.

2328F
Thiamine pyrophosphate deficiency secondary to TPK-1 mutation presenting as Leigh’s disease: diagnosis and management within a sibling pair. J.I. Fraser1, 2, S. Yang1, A. Vanderver4, T. Chang4, A. Lippe4, L. Cramp4, K. Chapman1, G. Vezina3, P. Šmpokou1, U. Lichter-Konečná1, D.J. Zandi1 1 Medical Genetics Branch/NHGRI/NIH, Bethesda, MD; 2 Department of Pediatrics, CNMC, Washington, D.C; 3 Division of Genetics and Metabolism, CNMC, Washington, D.C; 4 Department of Neurology, CNMC, Washington, D.C

Background: Thiamine pyrophosphate (TPP) is essential for pyruvate dehydrogenase complex, α-ketoglutarate dehydrogenase, and branched-chain ketoacid dehydrogenase function. Five patients have been reported with thiamine pyrophosphokinase-1 (TPK1) deficiency, with encephalopa-thy, extrapyramidal, bulbar, and cerebellar signs, and metabolic decomposition leading to death. Some patients received high dose thiamine therapy. Patients: We present a sibling pair with Leigh’s disease, progressive hypotonia, regression, and chronic encephalopathy. Whole exome sequencing in the younger sibling demonstrated homozygous TPK1 mutations. The older sibling had died from progressive neurologic disease with metabolic strokes, and the younger sibling had progressive neurologic decline. MRI demonstrated putamen and thalamic abnormalities in the younger, with similar and progressive evolution in the older sibling. The patient was started on high dose thiamine, niacin, biotin, and α-lipoic acid; ketogenic diet was initiated to reduce metabolic burden. He subsequently demonstrated improvement in neurologic function, with reattainment of lost milestones. Conclusions: TPK1 deficiency should be considered in children with Leigh’s disease without identifiable mutations, and this gene should be evaluated as a candidate for testing in commercially available sequencing panels. Reducing metabolic burden in these children by addition of supplements and transition to ketogenic diet may reverse some of the neurologic progression and improve outcome.

2329T
Deficiency of the mitochondrial phosphate carrier presenting with myopathy or cardiomyopathy: two new cases. L.C. Pyle, E.J. Bhog, M. Li, R.C. Ahrens-Nicklas, C. Ficicioglu, N. Sondheimer, M. Yudkoff. Section of Metabolic Disease, Children’s Hospital of Philadelphia, Philadelphia, PA. The SLC25A3 mitochondrial phosphate transporter is required to facilitate transport of inorganic phosphate across the inner mitochondrial membrane, for production of ATP via oxidative phosphorylation. Isoform A is expressed in heart and muscle, while B is the predominant isoform expressed in other tissues. Deficiencies in this transporter have been reported in five individuals to-date, across two sibships. Here we report two new cases of pathologic SLC25A3 mutations, in two different families, both presenting in the neonatal period. Case 1 presented as neonatal primary lactic acidosis, with subsequent improvement when intubated, but recurrent failure to extubate. Mitochondrional next generation sequencing revealed a homozygous IVS-9 mutation in SLC25A3, which has been previously shown in previous family 1 to cause a novel splice site, affecting an isoform specific to heart and skeletal muscle. This patient also has hypertrophic cardiomyopathy. Case 2 presented as neonatal severe hypertrophic cardiomyopathy. Mitochondrial next generation sequencing revealed two novel predicted pathogenic variants in SLC25A3. These cases expand our understanding of the SLC25A3 present-tation. They also highlight the relevance of mitochondrial studies in the setting of neonatal cardiomyopathy and neonatal primary lactic acidosis.
2330F
Use of next generation sequencing (NGS) in mitochondrial (mt) disorders: whole mitochondrial genome analysis. S. Seneca1,2, K. Vancampenhout2, G. Van Dyck2, J. Smets2, A. Van Lander3, R. Van Coster4, D. Danneels2, W. Lissens4,5, L. De Meirleir4. 1) Center for medical genetics, UZ Brussels, Brussels; 2) Vrije Universiteit Brussel (VUB, REGE), Brussels, Belgium; 3) University Hospital Ghent, Division of Paediatric Neu­rology and Metabolism, Ghent, Belgium; 4) UZ Brussels, Division of Paediatric Neurology, Brussels, Belgium.
Mitochondrial (mt) diseases are a heterogeneous group of disorders, caused by both nuclear and mt genome mutations. A correct diagnosis is challenging, mainly because of the absence of clear phenotype-genotype correlations, the existence of heteroplasmasy (presence of at least two different mt genotypes in the same cell) and the very large number of genes involved. Current traditional molecular diagnoses for disorders caused by a mt DNA defect relies on the identification of (common) point mutations and large deletions with sequencing and Southern blot procedures. Complete mt genome sequencing, using Sanger nucleotide sequencing techniques, is reserved only for few well selected patients. The method is laborious, and not very sensitive to detect nucleotide variations below 15-20% heteroplasmy. Next Generation Sequencing (NGS) is a booming technology, promising to be an accurate and cost effective method to investigate considerable amounts of DNA, including the complete mt genome of patients. DNA samples of thirty patients with (suspected) mt disorder were sequenced using the Ion PGM™ sequencer. Their mt genome has previously been characterized. A 100% coverage of the whole mt genome was obtained, 99.6% of the variants were concordant with Sanger sequencing. In depth sequencing allowed a sensitive detection of (pathogenic) heteroplasmy variations in a range of 4% to 79%, both in patients and controls. A sensitivity down to 4% is a major advantage in comparison with dideoxy nucleotide analysis. In addition, large multiple and single deletions, with visualization of their breakpoints, were identified. This study shows that NGS will be playing a major role in the molecular diagnostics of mt DNA disorders.

2331T
Purpose: Glucose transporter 1 deficiency syndrome (Glut1-DS) is a congenital metabolic disorder characterized by intractable seizures with early infantile onset, developmental delay, movement disorders and acquired microcephaly. The diagnosis is based on hypoglycorrhachia, impaired in vitro glucose uptake into erythrocytes, and heterozygous mutations in SLC2A1 gene. We report two patients with the same SLC2A1 mutation and different results in the erythrocyte glucose uptake to reveal clinical heterogeneity in Glut1-DS. Subjects and Methods: Subject 1 is a 15-year-old girl, born to healthy non-consanguineous parents. Atypical absence seizures started 6 years ago. Her glucose levels are normal, but hypoglycorrhachia and decreased erythrocyte glucose uptake were observed. She was diagnosed with Glut-1DS and started on the ketogenic diet. Subject 2 is a 17-year-old girl, born to healthy non-consanguineous parents. Since 4 years of age, she showed myoclonus, ataxia, and loss of consciousness with drooling. Such events were mostly seen in the late afternoon, and recovered with eating. Hypoglycorrhachia was observed while erythrocyte glucose uptake was normal. We did not start ketogenic diet, because Glut1-DS diagnosis was missed.
Results: We performed direct sequencing of DNA extracted from white blood cells of the two subjects, and identified the same mutation in SLC2A1 (R330X). To function of the R330X, we constructed wild type and R330X FLAG tagged SLC2A1 expression vectors and transfected to HEK293 cells. Wild type SLC2A1 expressed at plasma membrane while R330X stayed in cytoplasm. Thus SLC2A1 harbouring R330X is non-functional. Discussion: This is a first report of a Glut1-DS patient with normal erythrocyte glucose uptake caused by R330X mutation. There is evidence of heterogeneity in the level of erythrocyte glucose uptake in patients with the same SLC2A1 mutation, and it might be misleading. As Glut1-DS is a treatable epileptic encephalopathy, early diagnosis and treatment should be necessary. Conclusion: Glut1-DS caused by heterozygous mutations may result in various level of erythrocyte glucose uptake, thus mutation analysis of SLC2A1 should be performed for correct diagnosis.

2332F
Suspected mitochondrial myopathies in the era of genomic medicine. A.M. Atherton1,2, S.F. Kingsmore2, B.A. Hesse3,4, S.E. Soden3,4, C.J. Saun­ders5, E.G. Forrow6, L. Willig7, N.A. Miller7,12,13, I. Smith7,12,13. 1) Division of Genetics, Children's Mercy Hospitals and Clinics, Kansas City, MO; 2) Center for Pediatric Genomic Medicine, Children's Mercy Hospitals and Clinics, Kansas City, MO; 3) Pathology and Laboratory Medicine, Children's Mercy Hospitals and Clinics, Kansas City, MO; 4) Division of Developmental and Behavioral Sciences, Children's Mercy Hospitals and Clinics, Kansas City, MO; 5) Division of Pediatric Nephrology, Children's Mercy Hospitals and Clinics, Kansas City, MO.
Mitochondrial myopathies (MM), a heterogeneous group of inherited disorders with multisystem involvement, can be caused by genetic alterations in the nuclear or mitochondrial genome. Historically, individuals with a suspected MM were diagnosed based on blood and/or cerebral spinal fluid biochemical testing, neuroimaging or invasive muscle biopsy results. Non-specific results often lead to the diagnosis of a “suspected mitochondrial myopathy”. Extensive, generally expensive, molecular testing panels for genes causing MM are available but may also be uninformative. We explore the utility of exome sequencing for molecular diagnosis by analyzing whole exome sequences of 19 patients from 15 families with clinical symptoms suggestive of a mitochondrial disorder without a definitive diagnosis. Five of the patients (26%) were confirmed to have a primary MM, five patients (26%) were identified as having an unrelated disorder with secondary mitochondrial dysfunction, one patient (5%) has likely pathogenic variants in novel MM candidate genes and eight patients (42%) had negative testing results. Overall, 85% of patients received a molecular diagnosis utilizing whole genome sequencing. These results support the initial use of whole exome sequencing in patients suspected of having a MM as this testing modality is more practical, efficient, specific and cost effective.
Posters: Metabolic Disorders

2334F

Practical strategies for the identification of common mutations in Mucopolysaccharidosis IVA patients. F. Kubaski1, A.C. Brusius-Facchini2, P.F.V Medeiros3, R. Giugliani1,2, S. Leistner-Segal1,2,1. Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil; 2-Universidade Federal do Rio Grande do Sul/Post Graduation Program in Medicine: Medical Sciences; 3-Universidade Federal do Rio Grande do Sul/ Hospital Universitário Alcides Carneiro, Campina Grande, Brazil.

Mucopolysaccharidosis IVA or Morquio A syndrome, is an autosomal recessive disorder caused by deficiency of lysosomal enzyme N-acetylgalactosaminidase. These biochemical markers, the degree of enzyme deficiency, and the percentage of residual activity, can affect the diagnosis and treatment strategies. Here, we present a case report of a newborn from a consanguineous family, with a history of severe dysostosis of the newborn and, later, a coronal synostosis and a hypoplastic left heart. The child presented with delayed development, with ataxia, contractures, kyphoscoliosis, and joint laxity. At the age of 10 months, the child was able to sit with support, but was not able to roll over. At the age of 18 months, the child was able to sit, but was unable to crawl or walk. The child was referred to the geneticist at the age of 2 years. The patient was diagnosed with Morquio A syndrome, and was started on enzyme replacement therapy at the age of 3 years. The patient is currently able to sit, crawl, and walk with support. The patient's development is followed by the geneticist, and the patient is currently being treated with enzyme replacement therapy. The patient's family is being counseled regarding the patient's phenotype.

2336F

Biochemical, Molecular, and Clinical Presentations of 3 Patients with X-linked Epsilon-Trimethyllysine Hydrosylase (TMLHE) Deficiency, A Disorder of Carnitine Biosynthesis Associated with Autism Spectrum Disorder. M. Correia1, C. Schaff1, A. Petel1, Y. Yang1, F. Scaglia1, A. Beaudet1, S. Elsea1. 1-Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Previous research has shown that deletion of exon 2 in the gene encoding epsilon-trimethyllysine hydrolase (TMLHE) may be a risk factor for non-dysmorphic autism. Biochemical determination of trimethyllysine (TML) and gamma-butyrobetaine (GBB) in plasma and urine of affected individuals can be used to confirm the diagnosis of this disorder. There are three known phenotypes that present the same range of clinical, biochemical, molecular, and clinical phenotypes of three new patients with TMLHE deficiency. Patient 1 is a 4-year-old boy with a diagnosis of autism spectrum disorder who presented to genetics due to a recent episode of regression. Chromosome microarray (CMA) analysis was normal in this patient; however, whole exome sequencing found a hemizygous c.961_962del (p.I321fs) mutation in the TMLHE gene. Furthermore, biochemical analysis showed a low level of plasma carnitine, and analyze testing showed normal plasma and grossly abnormal urinary TML levels. The patient was subsequently placed on L-carnitine supplementation (200 mg/kg/day, divided in BID dosing), and his symptomatology has shown marked improvement in joint attention, eye contact, and social interaction. Patients 2 and 3 are twins born at 37 weeks gestation who presented at birth with congenital heart defects. CMA analysis found a 4.5 Mb deletion of 8p23.1, a region that contains over 20 genes and has previously been associated with congenital heart defects. Additionally, the CMA found a hemizygous deletion of Xq28 that encompassed exon 2 of the TMLHE gene with a minimum deletion size of Xq28:351145832_352168917. Sanger testing of plasma and urine revealed extremely high TML, low GBB, and normal carnitine values. These findings were discussed with the family and close monitoring of the twins’ development was recommended in light of the Xq28 finding. These cases highlight the benefits of biochemical testing in combination with CMA and whole exome studies toward early diagnoses and prevention of symptoms associated with autism. Furthermore, they emphasize the importance of including disorders of carnitine biosynthesis in the differential diagnosis of patients with autism spectrum disorder especially given that some of the associated phenotypes may be amenable to carnitine supplementation.

2337T

Detection of 4-hydroxy-2-oxoglutarate aldolase metabolites in urine for the diagnosis and monitoring of primary hyperoxaluria type III. L. Hasadrai, P. Loken, D. Gavrilov, D. Matern, K. Raymond, P. Rinaldo, S. Tortorelli, D. Oglesee. Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, MN.

Background: Mutations in the HOGA1 gene, encoding the mitochondrial enzyme 4-hydroxy-2-oxoglutarate aldolase (HOGA), were recently discovered to cause what is now known as primary hyperoxaluria type III (PH3). PH3 is a rare, autosomal recessive disorder characterized by the onset of recurrent kidney stone formation and progressive nephrocalcinosis during childhood. Molecular analysis of HOGA1 is currently available on a research basis, but the functional consequences of many variants found in this gene are still poorly understood. Timely diagnosis of patients with PH3 thus remains a significant challenge. HOGA catalyzes the conversion of 4-hydroxy-2-oxoglutarate (HOG) into glyoxylate and pyruvate, the final step in the metabolism of hydroxyproline. Patients with PH3 have been reported to have increased excretion of HOG and other HOG-related metabolites in their urine. Here, we describe the development of a novel method for the detection of urinary metabolites in PH3 via gas chromatography-mass spectrometry (GC-MS). Methods: Urine samples were treated with hydroxy-2-oxoglutarate at pH 9 and extracted with 5:1 (v/v) ethyl acetate/propan-2-ol, then evaporated to dryness under nitrogen. Residues were derivatized with BSTFA+TMCS and pyridine and analyzed on an Agilent 6890/5975 operating in scanning mode, with either helium or hydrogen as the carrier gas. Results: HOG was undetectable in urine by routine organic acid analysis, even when performed in splitless mode with high concentration of the analyte. When oxidized in an alkaline solution and extracted under more polar conditions, however, trimethylsilyl derivatives of HOG were easily identified, as were oxalic, glycolic, glyceral, glyoxylic, and pyruvic acid. Utilizing hydro-2-oxoglutaric acid as a standard, the metabolite run time was reduced by approximately 50%, and resulted in significantly improved resolution of the oxalate and glyoxylate peaks. Sample preparation time was <5 hours, and analytical run time was 8 minutes. Conclusions: GC-MS analysis of urine can provide a rapid, sensitive, and non-invasive method for detection and characterization of primary hyperoxaluria type III. Next steps include optimizing the assay’s extraction efficiency, and characterizing urine samples from individuals with a molecular diagnosis of PH3, asymptomatic heterozygous carriers, and patients with a strongly suspected diagnosis of PH3.
2338F
Unraveling mitochondrialopathies by exome sequencing. L.S. Kremer1, T.B. Haack1, R. Kopatch1, B. Haberberger1, C.A. Biagoshi3, T. Wieland1, T. Schwabe2, A. Althoff3, A. T.M. Strom3, J.A. Maurer1, W. Sport4, M. Zeviani1, P. Freisinger1, T. Klopotock5, R.W. Taylor1, A. Rötg1, A. Munich2, U. Ahting1, M.B. Hartig1, J.A.M. Smeitink6, G.F. Hoffmann6, A. Lombes1, T. Meitinger7, H. Proksch1. 1) Institute of Human Genetics, Universität Münster and Helmholtz Zentrum Münster, Germany; 2) MRC mitochondrial Biology Unit, Cambridge, United Kingdom; 3) Department of Genetics, Universiteit van Zuid Nederland, Maastricht, The Netherlands; 4) Department of Pediatrics, Paracelsus Medical University Salzburg, Salzburg, Austria; 5) Department of Neuropediatrics, Friedrich-Baur-Institute, Ludwig-Maximilians-Universität München, Munich, Germany; 6) Department of Pediatrics, University Hospital Heidelberg, Germany; 7) Department of Clinical Chemistry, Städtisches Klinikum München, Munich, Germany; 8) Nijmegen Centre for Mitochondrial Disorders, Department of Pediatrics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 9) INSERM, UMR 393 and Service de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 10) Mitochondrial Research Group, School of Neurology, Neurobiology and Psychiatry, The Medical School, University of Newcastle upon Tyne, United Kingdom; 11) INSERM, UMR 1016, Institut Cochin and AP-HP, Hôpital de La Salpêtrière, Service de Biochimie Météabolique et Centre de Génétique moléculaire et chromosomique, Paris, France.

Mitochondrial disorders are a genetically and clinically highly heterogeneous group of diseases characterized by faulty mitochondrial function. Despite progress in the field, most disease-causing mutations still have to be identified. We applied whole exome sequencing in 250 unrelated individuals with juvenile-onset mitochondrial disorder. In a quarter of patients, we detected mutations in known disease genes. In another quarter of patients, we identified mutations in genes previously not associated with mitochondrial disorders. While mutations in some genes are rare like MNGE1, the first exonic/mutation involved in mitochondrial replication (Kornblum et al., Nat. Genet. 2013), mutations in other genes are more frequent with ACAD9 being the most common finding with 13 cases. New examples are ELAC2, a 2340F
Creative Deficiency Due To Targeted Disruption Of Alanine: Glycine Amidotransferase Leads To Learning And Memory Deficits. W. Craigen1,2, Y. Le1, M. Costa-Mattolli3, L. Stoica2, P.J. Zhu1, L. Mye1. 1) Molecular & Human Genetics; 2) Pediatrics; 3) Neuroscience, Baylor College of Medicine, Houston, TX.

Phosphocreatine serves as an important reservoir of high-energy phosphate for ATP synthesis in tissues that have fluctuating demands for energy. Human disorders of creatine biosynthesis and transport exist, leading to intellectual disabilities, epilepsy, and poor growth. Creatine biosynthesis requires two enzymes, the first, L-arginine:glycine amidotransferase, the product of the Gatm gene, catalyzes the rate-limiting transfer of the amino group from arginine to glycine, yielding guanidoacetate and ornithine. We generated a mouse strain lacking Gatm. Homozygous Gatm deficient mice have an almost complete absence of creatine and guanidoacetate, and exhibit poor growth and muscle weakness. A consequence of creatine deficiency is the global reduction in skeletal muscle and brain mitochondrial respiratory chain activities, and a compensatory increase in mtDNA content and citrate synthase activity. Gene expression profiling reveals the activation of the integrated stress response. When challenged with the seizue-inducing drug kainate, creatine deficient mice are resistant to the induction of status epilepticus and epilepsy-associated mortality. When tested for associative learning by fear conditioning paradigms, creatine deficient mice have a marked deficit in learning, and are similarly compromised in the Morris water maze test. This animal model of creatine deficiency demonstrates the central role creatine metabolism plays in normal neurologic functions.
2342F

The clinical and mutation spectrum of Korean patients with urea cycle disorders. H.W. Yoo1,2, B.H. Lee1,2, G.H. Kim3, J.M. Kim2, Y.M. Kim1, J.H. Kim1.

1) Department of Pediatrics, Asan Medical Ctr, University of Ulsan College of Medicine, Seoul, Korea; 2) Medical Genetics Center, Asan Medical Ctr, Korea.

Urea cycle disorder (UCD) is caused by the deficiencies of the six enzymes, carbamoyl phosphate synthetase (CPS), N-acetyl glutamate synthetase (NAGS), ornhime transcarbamoylase (OTC), argininosuccinate synthetase (ASS), argininosuccinate lyase (ASL) and arginase (ARG) or mitochondrial ornithine transporter (ORN1). This study was undertaken to review the clinical and molecular genetic characteristics of Korean UCD patients diagnosed in one institution. A total of 68 patients, 30 males and 38 females, from 66 unrelated families were diagnosed based on genetic and biochemical profiles during past 15 yrs. UCD manifested in neonatal period in 38 patients (64.4%), whereas 27 patients (45.8%) were late onset UCD. The remaining two patients (3.4%) were identified in presymptomatic periods. Thirty patients (44.1%) were diagnosed as OTC deficiency, 22 (32.4%) as ASS deficiency, 10 (14.7%) as CPS deficiency, 3 (4.4%) as ASL deficiency, 1 as argininaemia, 1 as Hyperornithinemia-Hyperammonemia-Nonketotic hyperglycinemia (HHH) syndrome, and 1 as NAGS deficiency. Mutation detection rate was 97.7% (43/44 alleles) in the ASS1 gene, 90.0% (24/27 alleles) in the OTC gene, and 100% (6/6 alleles) in the ASL gene, whereas it was 55.6%(10/18 alleles) in the CPS1 gene. Total 25 novel mutations were identified. Most patients presented as symptomatic hyperammonemia, whereas an argininaemia patient as spasitic diplegia without pronounced hyperammonemia, and 2 patients were identified in asymptomatic period by familial screening or tandem mass screening. Among 42 patients with hyperammonemia >300umol/l, 25 were managed by peritoneal dialysis or continuous renal replacement therapy, whereas 17 managed conservatively. The clinical outcomes were normal neurological development (19 patients, 32.2%), neurological deficit (26 patients, 44.1%), and death from hyperammonemia (13 patients, 21%). The mutational analysis of the genes was performed in 3 patients. This study portrayed the full spectrum of clinical and molecular genetic characteristics of UCDs in Korea. More efforts are needed for early identification and urgent management of each patient with UCD to improve the long-term neurological outcome.

2343T

Leigh syndrome and myoclonic epilepsy caused by novel mutation in AIFM1 gene. T. Honzik1, A. Vondrackova1, V. Strancky2, M. Rodinova2, H. Kratochvilova1, H. Harskova2, P. Klement2, M. Magner1, S. Mazurova1, J. Zeman1, M. Tesarova1.

1) Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Czech Republic; 2) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Czech Republic.

Mutations in X-linked AIFM1 gene, encoding the Apoptosis-Inducing Factor Mitochondrion-associated 1, cause disruption of the respiratory chain complexes, decrease stability of mtDNA and increase apoptotic stimuli. Only 11 patients from 3 families were reported, so far. Severe encephalopathy and Leigh syndrome was present in four children and progressive axonal neuropathy, deafness and cognitive impairment were present in 7 patients (Cowchock syndrome; CMTX4). Here we describe another patient with neonatal onset of the disease with hypotonia, myoclonic epilepsy, hearing loss, visual impairment, severe encephalopathy, Leigh syndrome and lactic acidosis. In addition, the boy had hyporeflexia compatible with axonal neuropathy and died at the age of 18 months. Muscle biopsy demonstrated mild myopathic pattern with accumulation of SDH product. In tissues obtained at autopsy, the activity and amount of complex I and IV was decreased in muscle, brain and heart and the activity of complex II was upregulated. Targeted sequencing of mitochondrial exome revealed novel hemizygous mutation c.1391T>G (p.Leu464Trp) in AIFM1 gene on chromosome X. Conclusion: X-linked AIFM1 encephalomyopathy due to apoptosis-inducing factor defect should be considered in the differential diagnostics in patients with Leigh syndrome, axonal neuropathy and myoclonic epilepsy. Supported by IGA NT 14156/3 and IGA NT 13114/4.

2344F


1) Molecular Medicine Program; NICH/NIH, Bethesda, MD; 2) Undiagnosed Disease Program; NIH/NIH, Bethesda, MD; 3) Clinical Neurosciences Program; NINDS, NIH, Bethesda, MD.

We evaluated a 6 year old white male from the NIH Undiagnosed Disease Program for a possible Menkes disease variant. His term birth was normal with Apgar scores recorded as 8 at both 1 and 5 minutes. Early neurodevelopment was markedly delayed. Since infancy, he has remained a diagnostic conundrum despite extensive testing. Based on normal plasma neurochemicals at age two years, Menkes disease was excluded. Physical examination findings at age 6 years included dysmorphic facies, coarse hair texture, alternating esotropia, generalized hypotonia, mild joint laxity, and moderate intellectual disability. Hair analysis showed pili torti in 4/25 (16%) of hair shafts obtained at age 9 months, and 7/35 (20%) hair shafts at age 6 years. A positive control hair specimen, obtained from a 3 year old classic Menkes patient receiving copper injection treatments, showed 6/25 (24%) pili torti. A brain MRI/MRA at age 6 years showed vascular dilatation and tortuosity of the major arteries within the Circle of Willis. The patient had low-normal levels of serum copper (64 µg/dl) and ceruloplasmin (172 mg/L). Plasma catechol ratios were normal when tested at age 2 years and again at age 6 years, indicating normal peripheral activity of dopamine-beta-hydroxylase (DBH), a copper-dependent enzyme. However the catechol ratios in his cerebrospinal fluid at age 6 years were just within the range from a cohort of classic Menkes disease patients. DNA sequence analysis failed to reveal any abnormality in the ATPTA coding regions, splice junctions, and known regulatory elements. However, Western blot assay of the patient's fibroblasts revealed approximately 90% reduction in ATPTA protein quantity compared to wild-type cells. The residual ATPTA protein was normal-sized and exhibited proper intracellular trafficking in vitro in response to varying concentrations of copper. The patient's phenotype does not fit well with any of the three recently identified autosomal recessive copper metabolism phenotypes: (1) CNS obesity with mutations disrupting ABC transporters (ATP7A, ATP7B) transporter (AT-1), or 3) the sigma 1 alpha subunit of adaptor protein complex 1 (AP1S1). We therefore hypothesize a defect in an unrecognized transcriptional regulator of ATPTA and predict that delineation of the precise transcriptional role(s) of this syndrome (AIFM1), will reveal yet another novel cause of abnormal human copper metabolism, and one which may be amenable to therapeutic intervention.

2345T


Center For Gene Protein research, Hanoi Medical University, Hanoi, Vietnam.

Wilson disease (WD) is an autosomal recessive inherited disease caused by abnormalities of the copper-transporting protein encoding gene ATPTB with over 600 mutations described. Identification of mutations has made genetic diagnosis of WD feasible in many countries. In this study, we examined ATPTB for mutations in 19 unrelated Vietnamese WD patients who were diagnosed as WD base on typical clinical symptoms. Their clinical and laboratory findings were recorded. DNA sequencing of the ATPTB coding regions, splice junctions, and known regulatory elements revealed approximately 90% reduction in ATPTA protein quantity compared to wild-type cells. The residual ATPTA protein was normal-sized and exhibited proper intracellular trafficking in vitro in response to varying concentrations of copper. The patient's phenotype does not fit well with any of the three recently identified autosomal recessive copper metabolism phenotypes: (1) CNS obesity with mutations disrupting ABC transporters (ATP7A, ATP7B) transporter (AT-1), or 3) the sigma 1 alpha subunit of adaptor protein complex 1 (AP1S1). We therefore hypothesize a defect in an unrecognized transcriptional regulator of ATPTA and predict that delineation of the precise transcriptional role(s) of this syndrome (AIFM1), will reveal yet another novel cause of abnormal human copper metabolism, and one which may be amenable to therapeutic intervention.
2346F


NGLY1 encodes N-glycanase, a deglycosylating enzyme, which is proposed to be a component of the ER-associated degradation (ERAD) pathway. This pathway is responsible for the degradation of misfolded glycoproteins. Defects in NGLY1 are predicted to result in the accumulation of misfolded glycoproteins due to impaired degradation. We performed clinical whole exome sequencing on a 2 year-old girl and identified a homozygous nonsense mutation c.1201A>T (p.R401X) in exon eight of the NGLY1 gene. Both parents are carriers of this mutation. An older sister with a similar phenotype also has the homozygous nonsense mutation. Two unaffected siblings do not carry this mutation. The clinical features of the 2 yr old girl and her sister include severe intellectual disability, abnormal movements, absent speech, hypotonia and mild dysmorphic features. She had an extensive negative evaluation including: a chromosomal microarray, plasma amino acid, urine organic acid and acyl carnitine analysis, transferrin and N-glycan analysis, brain MRI, muscle biopsy, mitochondrial genome sequencing, lysosomal enzyme panel, and MECPP2 analysis. The p.R401X mutation has been recently reported in another individual in trans with a c.1891delC mutation. Similar phenotypic features shared among all three individuals include developmental delay and involuntary movements. The movement disorder in our patient and her sister were notable for occasional hand-wringer reminiscent of classic Rett syndrome. We are currently screening MEOP2 negative patients as well as children with similar phenotypes for NGLY1 to determine the prevalence of this novel neurogenetic disease. We propose that NGLY1 should be considered in the differential diagnosis for patients with severe intellectual disability and abnormal movements. Identification of additional patients with this rare disorder will allow further characterization of the phenotypic spectrum and natural progression of the disease.

2347T


As an attempt to predict the functional consequence of substituted nucleotides in the ATP7A gene we have investigated the effect of 36 selected missense mutations in the ATP7A gene according to the resulting ATP7A transcript and protein including the intracellular localization. ATP7A encodes a copper-translocating ATPase belonging to the family of P-type ATPases. Mutations of ATP7A lead to Menkes disease, which is an X-linked, lethal disorder of copper metabolism. The ATP7A protein has a dual function: in basal copper levels ATP7A is located in the trans-Golgi network where it is responsible for delivery of copper across the membrane for cuproenzyme biogenesis. In excess intracellular copper levels, the steady state distribution of ATP7A shifts to the plasma membrane where the protein is responsible for pumping copper from the cell. Fibroblasts obtained from Menkes patients with missense mutations in the ATP7A gene were used as source for ATP7A mRNA and protein investigation. Investigation of ATP7A transcript was performed by RT-PCR and real-time PCR. ATP7A Protein quantitation and localization revealed was performed by Western blotting and immunofluorescence. Notably, the pathogenic effect of almost one third of the missense mutations is not due to functional importance of the affected residue, as the mutations affect splice sites, splicing silencers, mRNA stability or protein stability, leading to absence or reduction of the resulting ATP7A protein. More than two third of the mutations abrogates copper dependent trafficking, probably by affecting phosphorylation, dephosphorylation, copper binding or copper transport. Only two out of 36 missense mutations show normal copper dependent trafficking. These results emphasized the importance of using a comprehensive spectrum of analyses to reveal the deleterious effect of missense mutations.

2348F

The effect of homozygous deletion of BBOX1 gene on carnitine level and fatty acid beta-oxidation. A. Rashidi-Nezhad1,2, S. Talebi2, S.M. Akrami2, A. Reymond2. 1) Maternal, Fetal & Neonatal Research Center, Tehran University of Medical sciences, Tehran, Iran; 2) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 3) Department of Medical Genetics, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.

Carnitine is a key molecule in energy metabolism and functions in the transport of activated fatty acids into the mitochondria. Carnitine homestasis achieved and remained through oral intake, renal reabsorption and de novo biosynthesis. Unlike the dietary intake and renal reabsorption, the importance of de novo biosynthesis pathway in carnitine homestasis remains unclear, due to lack of animal models or patients with the defect of this pathway. Here we evaluate metabolic and molecular study on carnitine level in a patient deficient in carnitine de novo biosynthesis. The proband was a 42 months old girl with microcephaly, speech delay, growth retardation and minor facial anomalies. ArrayCGH was used to detect small genomic imbalances. Quantitative PCR (qPCR) was used to confirm array results. Carnitine profile and metabolic investigation were performed according to routine laboratory protocols. The proband showed 251 Kb interstitial homozygous deletions at 11p14.2. Since BBOX1 gene is located in this region, the proband has no de novo biosynthesis of carnitine. Although free carnitine level in the proband was in normal range, acylcarnitine to free carnitine ratio (AC/FC ratio) was high but near the upper limit of reference range. Almost all the other evaluated metabolites were normal. In conclusion, we present the first report of a patient with the complete defect of carnitine de novo biosynthesis. Although this condition can result in a slight increase in AC/FC ratio, it does not give rise to clinical manifestations of carnitine deficiency disorders.

2349T

Glycerol Kinase Transgenic Mice Have Increased Risk For Obesity. C. Ho1, A. Badjatiya1, K.M. Dipple2,3. 1) Dept of Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Dept of Pediatrics, Division of Genetics, University of California. Los Angeles, CA.

Glycerol kinase deficiency (GKD) is an X-linked inborn error of metabolism with a lack of genotype-phenotype correlation. Glycerol kinase (GK in humans, Gyk in mice) not only phosphorylates glycerol, but also performs other protein functions that may explain the pathogenesis of GKD. We recently developed a liver-specific Gyk transgenic mouse strains to understand the alternative functions of GK, including its role in adipogenesis and obesity. Male wildtype (WT/WT), heterozygous (WT/T) and homozygous (T/T) transgenic mice were placed on either regular chow or high fat (HF) diet for 12 weeks and monitored for weight gain and percentage body fat. Fasting glucose and cholesterol levels were measured from blood plasma to assess their risk for obesity and type II diabetes mellitus (T2DM). WT/T and T/T mice on HF diet gained more weight, 2.8% and 11.3% respectively, than WT/WT mice (p<0.05). Nuclear magnetic resonance analysis revealed both wildtype and transgenic mice had at least two-fold increase in percentage body fat as a result of consuming the HF diet. The elevated body weight in transgenic mice was also validated based on weighing of various tissues, including the mass of liver, visceral fat, perirenal fat and subcutaneous fat pads. T/T mice on chow diet had significantly greater mass in the liver tissue (25.8%), visceral fat pads (53.3%), and perirenal fat pads (40.2%) compared to the WT/WT mice (p<0.05), suggesting that Gyk overexpression is associated with abdominal obesity. Blood glucose and cholesterol tests showed that T/T mice on HF diet had higher triglycerides levels (16.4%), total cholesterol levels (41.3%), and fasting glucose levels (45.4%) relative to WT/WT, also implying high risk for obesity and T2DM (p<0.05). This study demonstrates that GK is involved in fat deposition and adipogenesis, and its overexpression increases risk of obesity and T2DM in mice. In addition, we propose that our Gyk transgenic mouse strains may serve as a useful model for obesity and T2DM research.
Novel heterozygous and homozygous mutations in the gap junction protein gamma-2 (GJC2) gene cause Pelizaeus-Merzbacher like disease (PMLD), a rare hypomyelinating leukodystrophy. The clinical phenotype of PMLD is similar to that of the X-linked Pelizaeus-Merzbacher disease (PMD) and includes features such as spasticity, ataxia, dystonia, nystagmus, and developmental delay. A distinguishing feature of PMLD is the presence of brainstem auditory evoked potentials, which are typically absent in PMD. There have been reports that dominant point mutations in GJC2 are associated with hereditary lymphoedema type IC. Associated features include mild to severe swelling of the limbs and recurrent cellulitis. For the first time in our Molecular Diagnostics Lab, we report a cohort of patients and family members who have mutations in GJC2. For each patient, the entire coding region, surrounding intron, and promoter region of GJC2 were sequenced. Family members, when available, were screened for the mutation(s) seen in the proband. Any changes from the reference sequence were analyzed via SIFT, PolyPhen, and Mutation Taster. These mutations include novel and previously reported single heterozygous, homozygous, and compound heterozygous mutations.

The mutations are scattered throughout the entire coding region, as well as two different mutations in the promoter region. Three patients from different families had the c.-167A>G promoter mutation, which supports the previously reported hypothesis that this is a common pathogenic mutation in GJC2. The mutations p.Pro90Ser, p.Arg104Val*106, p.His252Arg, and p.Thr398Ile were each seen in two different families. Both the p.Pro90Ser and p.Arg104Val*106 mutations occur as homozygous traits in the affected patients. However, these two mutations are also relatively common. For those patients with only one heterozygous mutation in the GJC2 gene having PMLD symptoms, we suggest the possibility of an additional mutation being present in a region not sequenced, variable expressivity, or autosomal dominant inheritance. Further parental and familial genetic testing and phenotypic review should be completed to help make this distinction.

Arylsulfatase A Deficiency in Thai MLD Patients.

Metachromatic Leukodystrophy (MLD) is the lack of Arulsulfatase A (ARSA) enzyme which missing of ARSA enzyme lead to damage nervous system, kidney and other organs. MLD is an autosomal recessive disease caused by the absence of ARSA enzyme which is encoded by the ARSA gene. ARSA is a lysosomal enzyme that catalyzes the hydrolysis of the glycosphingolipid sulfatides. The deficiency of ARSA leads to the accumulation of these substrates in the acidic vesicles of the lysosomes, resulting in the characteristic white matter changes of the brain. MLD is classified into two main forms: adult and juvenile. Adult MLD is characterized by early-onset cerebellar ataxia and spastic paraplegia, while juvenile MLD presents with more severe neurological symptoms, including cognitive impairment, spastic paraplegia, and dystonia.

In our study, we sequenced the ARSA gene in a cohort of Thai MLD patients. We identified two novel mutations in the ARSA gene: c.214C>T and c.619A>G. These mutations were found in two unrelated families. The c.214C>T mutation results in a Leu->Phe substitution at codon 72, while the c.619A>G mutation leads to a Thr->Ser substitution at codon 207. The clinical presentation of these two patients was consistent with juvenile MLD, characterized by severe neurological symptoms and cognitive impairment.

Comprehensive mutation screening of the AGXT gene in patients with primary hyperoxaluria type 1.

Primary hyperoxaluria type 1 (PH1) is a rare autosomal recessive disorder caused by deficiency of alanine glyoxylate aminotransferase (AGT) enzyme encoded by the AGT gene. It is characterized by recurrent nephrolithiasis, nephrocalcinosis, and end stage renal disease. The three most common mutations in the exons 1, 4, and 7 have been found in more than 50% of the cases. Hence, mutation screening of these three exons has suggested being the first-line test to avoid liver biopsy. Since the facilities for liver biopsy are not commonly available in this region, mutation screening is the only reliable procedure for the diagnosis of PH1. Mutation analysis was carried out by direct sequencing of the AGXT gene in 150 patients. The age of onset of PH1 was 1-16yrs. They rapidly progressed to renal failure at adolescence. The study was approved by the IRB and conformed to the Tenets of the Declaration of Helsinki. We have found 17 mutations in 44 patients in which three mutations were novel. They included one splice-site, four indel and twelve missense mutations. The most common p.G350D mutation was found in 12 patients (8%). The age of onset of PH1 in these patients was 5-16 yrs. Of these, three children developed renal failure. One patient had encountered transplant rejection due to the recurrence of oxalosis. Another common mutation p.G190R was found in 7 cases (4.6%). Three patients were on the background of the AGT-Ma allele. Interestingly, the most commonly reported mutation c.508G>A (p.G170R; 20-40% allele frequency) was not found in our patients. Another common c.33dupC (12% allele frequency in Caucasian patients) that lead to protein truncation was found in only five cases (3.3%). AGT-Ma allele was found in 12 patients (8%) that mostly occurred in combination with the p.G350D mutation which is not reported elsewhere. Linkage disequilibrium among the AGT-Ma was not found in this study. Mutations in the AGXT gene have been identified in more than 90% of the cases worldwide. However, they were found in only 30% of the cases in this cohort. It is noteworthy that the most commonly reported mutations are not prevailing in our patients, therefore, mutation screening of the whole AGXT gene needed to be carried out in these patients.
2354F
Mutation of the iron-sulfur cluster assembly IBAS7 gene causes lethal myopathy and encephalopathy. N. Aijl Bolar, A. V. Vanlanderde, N. Van den Oetel, B. De Paepen, G. Vandeweyer, E. Kooy, E. Eyskins, E. De Lattier, G. Delanghe, P. Govaert, J. G. Leroy, R. Lig, R. Van Coster, L. Van Laer, B. Loeyes. 1) Department of Medical Genetics, University of Antwerp, Antwerp, Edegem, Antwerp, Belgium; 2) Department of Pediatrics, Division of Pediatric Neurology and Metabolism, University Hospital Ghent, 9000 Ghent, Belgium; 3) Institute for Virology, Philippus-Universität Marburg, Robert-Koch Str. 6, 35032 Marburg, Germany; 4) Provincial Centre for the Opsporing van Metabolie Aandoeningen (OPMA) at the Department of Pediatrics, University Medical Centre Utrecht, Utrecht, The Netherlands; 5) Department of Medicine and Health Sciences, Antwerp University Hospital and University of Antwerp, 2000 Antwerp, Belgium; 6) Department of Neonatology, Paola Children’s Hospital ZNA Middelheim, 2000 Antwerp, Belgium; 7) LOEWE Zentrum für Synthetische Mikrobiologie SynMikro, Hans-Meerwein-Str., 35043 Marburg, Germany; 8) The Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Office of the Director, National Institutes of Health, Bethesda, MD. Cobalamin C disease (cblC) is the most common inborn error of intracellular cobalamin metabolism. It is caused by mutations in the MMACHC, a gene responsible for processing and trafficking intracellular cobalamin. Defects in this pathway impair the function of two cobalamin-dependent enzymes: methylmalonyl-CoA mutase and methionine synthase. Disease manifestation is severe, showing neurological problems, growth retardation, heart defects, and progressive blindness. At present, the pathological basis of these symptoms remains unknown, and no animal model exists. To replicate clinical manifestations experienced by patients with cblC disease, we created a series of loss of function alleles in the zebrafish ortholog of MMACHC using zinc-finger nucleases. Of these, we chose p.L44PfsX21 (hg12) and p.G32VfsX48 (hg13), transmitted by two independent founders, for phenotype analysis. F2 mmachc<sup>hg12/hg12</sup> and mmachc<sup>hg32/hg32</sup> fish survived the embryonic period but displayed growth impairment after 14 days post-fertilization (dpf). By 21 dpf, the standard length (SL) and height at the anterior of the anal fin (HAA) were significantly reduced; mmachc<sup>hg12/hg12</sup> fish (SL 6.94 ± 0.07, HAA 0.77 ± 0.03 mm) and mmachc<sup>hg32/hg32</sup> (SL 7.40 ± 0.07, HAA 0.86 ± 0.01 mm) fish were smaller than wild-type and heterozygous fish (SL 10.39 ± 0.18, HAA 1.48 ± 0.03 mm) (p < 0.0001). Histological examination of mmachc<sup>hg12/hg12</sup> fish revealed a complete absence of the secondary lamellae in the gills, which contain specialized cells for gas and ion exchange. Thinner retinal layers and a possible defect in the morphology of the retinal pigmented epithelium were also detected. The zebrafish model of methylmalonic acid (MMA), a classic biomarker of cblC disease, was elevated by 289-fold in mmachc<sup>hg12/hg12</sup> fish. OH-cobalamin (OH-cbl) injections are the main treatment administered to the patients and ameliorate metabolic symptoms in cblC patients. When fish were maintained in water supplemented with OH-cbl (100 µg/ml) for 21 days, SL increased by 25% (p<0.05) and HAA increased by 30% (p<0.01) compared to the untreated group. The zebrafish model of cblC disease we generated recapitulates several of the phenotypic and biochemical features of MMACHC deficiency, demonstrates a response to conventional therapy, and should be useful to delineate the pathophysiological mechanisms in this common disorder of cobalamin metabolism.

2357T
Identification of a novel mutation in the human ARSB gene on chromosome 5q14.1 for Muchopolysaccharidosis type VI patients in southwest Colombia. M.A. Acosta Aragón, J.R. Lago, F. Barros, A.M. Carracedo Alvarado, N.P. Achilly, R.B. Sood, K.S. Bishop, M.S. Jones, V.J. Hoffman, R.B. Song, G.P. Ventitti. 1) Human Genetics Program, NHGRI, 5600 Rockville Pike, MSC 6225, Bethesda, MD 20852, USA; 2) Instituto de Medicina Tropical “Dr. Caldas” Bogota, Bogotá, Colombia; 3) Departamento de Médica Xeral, Facultade de Medicina, Universidade de Santiago de Compostela, Galicia, Spain; 4) Department of Paediatrics, University of Antwerp, Wilrijk, Belgium; 5) Laboratorio de Batismo e Metabolismo, Hospital Universitario de Getafe, Madrid, Spain; 6) Servicio de Medicina Genética, Hospital Universitario de Getafe, Madrid, Spain; 7) LOEWE Zentrum für Synthetische Mikrobiologie SynMikro, Hans-Meerwein-Str., 35043 Marburg, Germany; 8) The Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Office of the Director, National Institutes of Health, Bethesda, MD. Muchin protein (MUC) and ganglioside binding activity (GBA) are the main treatment administered to the patients and ameliorate metabolic symptoms in cblC patients. When fish were maintained in water supplemented with OH-cbl (100 µg/ml) for 21 days, SL increased by 25% (p<0.05) and HAA increased by 30% (p<0.01) compared to the untreated group. The zebrafish model of cblC disease we generated recapitulates several of the phenotypic and biochemical features of MMACHC deficiency, demonstrates a response to conventional therapy, and should be useful to delineate the pathophysiological mechanisms in this common disorder of cobalamin metabolism.

2355T
Genetic variation in a gene involved in glycosphingolipid biosynthesis. O.M. Amaral, A.J. Coelho, E. Pinto, R. Lopes, R. Ribeiro. Department of Human Genetics, INSA, IP., 4000-Porto, Portugal. The main objective of this work was to investigate the possible existence of genetic variation in the UGCG gene. The UGCG gene encodes an enzyme (UDP-N-acetylgalactosamine 4-sulfatase) responsible for catalyzing the formation of sulfated glycosaminoglycans. The genetic variation could lead to differences in biosynthesis and be related to phenotypic divergence in various genetic diseases of the glycosphingolipidoses group. In order to test this hypothesis we attempted to identify the extent of variation in the UGCG gene in order to relate it to phenotypic variation. Methods and samples: DNA was extracted from blood samples and/or fibroblast cell lines using an automated apparatus. Biological samples were obtained from healthy donors, with informed consent. In addition, all traceable identification was removed, so as to guarantee their anonymity. Skin fibroblast cell lines were obtained from the Corriel Institute (USA). The UGCG gene (exons and flanking intronic regions) of six control individuals was sequenced using standard methods. Results: In this work we present the identification and distribution of genetic variations among the six control individuals. The results obtained by the different laboratories showed the existence of several polymorphic changes. Discussion: Polymorphisms in the UGCG gene may interfere with the amount of substrate available for degradation in specific diseases along the same pathway. Thus, the degree of genetic variability might influence the phenotypic expression as well as the lysosomal burden. Conclusion: Assessment of variation in the UGCG gene should be considered, particularly in patients who do not comply with the expected genotype/phenotype correlations. Additional information: This work was supported by the Fundação para a Ciência e Tecnologia (CT/MCTES) - Portugal. Corresponding author: Olga Amaral, olga.amaral@insa.min-saude.pt.
Two novel mutations in α-glucosidase gene in two patients with Pompe disease. A. Aykul1, H. Onay2, M. Kose3, E. Erbas Canda1, E. Karaca1, M. Coker2, F. Ozkinyaz1. 1) Department of Medical Genetics, Ege University Faculty of Medicine, Izmir, Bornova Izmır, Turkey; 2) Department of Pediatric Metabolic Disease, University Medical Faculty, Izmir, Turkey.

Glycogen storage disease type II (GSD-II), also known as Pompe disease (OMIM #232600), is a rare progressive autosomal recessive disorder characterized by deficiency of α-glucosidase (GAA) that results in abnormal glycogen deposition in the muscles. Mutations in the GAA gene cause Pompe disease. The infantile-onset form is the most severe form. Most of the infantile-onset form patients present with hypotonia and cardiomyopathy in early infancy. The first case presented is a second child of consanguineous parents, born at term after an uneventful pregnancy. At the age of 9 months infantile type Pompe disease was suspected based on hypertrophic cardiomyopathy. Unfortunately her previous sister died from hypertrophic cardiomyopathy suggesting Pompe disease at 7 months old. The diagnosis was confirmed by measuring α-glucosidase activity in serum (0.4 µkat/kg protein). Although enzyme replacement therapy was started at 9 months, patient died from sudden cardiac arrest at 1 year old. Genetic analysis of the mother revealed GAA enzyme activity that was less than 1 percent of normal. She died from hypertrophic cardiomyopathy at 2 years old. Genetic analysis of the father revealed a novel missense mutation p.Q682R in GAA gene which was not previously reported. Second case, is a child of nonconsanguineous parents, born at term after an uneventful pregnancy. Infantile type Pompe disease was suspected based on hypotonia/generalized muscle weakness at 6 months. Sensorineural hearing loss and optic atrophy were detected. Echocardiography was normal. Levels of creatine kinase, or creatine kinase-myocardial band isoenzyme were typically elevated. She had respiratory weakness at 6 months. Sensorineural hearing loss and optic atrophy were detected. Echocardiography was normal. Levels of creatine kinase, or creatine kinase-myocardial band isoenzyme were typically elevated. She had respiratory weakness at 2 years old. Genetic analysis of the mother revealed a novel nonsense mutation p.Q255X in GAA gene which was not previously reported and father had p.R224W which was previously described. We report two families wherein molecular diagnosis for Pompe disease was performed and we have identified two novel mutations in the GAA gene which were not previously described.

New dominant ALDH18A1 mutation in two unrelated children with neurodevelopmental delay, cataracts, cutis laxa, and intracranial arterial tortuosity. J. Ganesh1, A.E. Lin2, I. Sahai2, E.H. Zakkail1, S. Chaddock1, T.L. Toler2, P.H. Byers3, A. Schwarze3. 1) Children’s Hospital of Philadelphia, Philadelphia, PA; 2) Mass General Hospital for Children, Boston, MA; 3) Department of Pathology, University of Washington, Seattle, WA; 4) Department of Medicine, University of Washington, Seattle, WA.

Cutis laxa syndromes are genetically heterogeneous. To date, mutations in ALDH18A1, encoding Δ1-pyruvate-5-carboxylate synthase (P5CS), have been reported in five families. In four families biallelic mutations were present, and in the fifth there was a causative heterozygous mutation. We identified an apparently dominant ALDH18A1 mutation (c.413G>A, p.Arg138Gln) in two unrelated infants with a complex phenotype that overlapped with the findings in reported patients: delayed development, failure to thrive, neurodevelopmental delay, bilateral cataract, intracranial arterial tortuosity, skin laxity, and a metabolic profile of low serum ornithine, citrulline, arginine, proline, and mildly elevated ammonia. P5CS is a bifunctional mitochondrial enzyme with both gamma-glutamyl kinase (γ-GK) and gamma-glutamyl phosphate reductase activity. It catalyzes a critical step in the biosynthesis of proline and ornithine, a substrate of the urea cycle and precursor in arginine biosynthesis. The heterozygous mutation that we identified in both infants results in substitution of an evolutionarily conserved arginine by glutamine at position 138 within the γ-GK domain. In the first child (P1) the mutation occurred de novo on the paternally derived allele. In addition there was a maternally derived polymorphic variant (c.896C>T, p.Thr299Ile) that has a heterozygote frequency of about 20% among European population. In the second child (P2) the mutation arose de novo and there was no background polymorphic minor allele. We identified no other alteration in the gene sequence or relative mRNA abundance of the products of the two alleles in fibroblasts. These findings are consistent with the concept that the substitution at this site represents a new dominant mutation and is sufficient to cause the phenotype. P1 was treated with L-arginine, leading to resolution of irritability, feeding difficulties, and tremors within a few weeks, and sustained improvement in growth and development after six months of treatment. Intracranial arterial tortuosity was identified by ultrasound and was stable on serial magnetic resonance imaging. In the growing number of other “cutis laxa genes” has not led to a diagnosis.

Molecular Characterisation of known mutations in Congenital Adrenal Hyperplasia patients (CYP21A2 gene) : Genetic & Diagnostic Implications. R. Khajuria1, A. Kumar1, D. Pal1, U. Sharma2, A. Bhanushali2, R. Waliya2, R. Prasad1. 1) biochemistry, PGIMER, CHANDIGARH, India; 2) ENDOCRINOLOGY, PGIMER, CHANDIGARH, INDIA.

Abbreviations : SW=Salt Wasting, SV=Simple Virilising, NC= Non classical, ACRS-Amplification created restriction site. Background : Congenital Adrenal Hyperplasia (CAH) is frequent autosomal recessive disease, with a wide range of clinical manifestations from severe classical form to late onset form. Most cases of CAH, the inherited inability to synthesize cortisol, aldosterone with subsequent overproduction of androgens, are caused by mutations in steroid 21 Hydroxylase gene. As a complement to hormonal measurements, mutation analysis of CYP21A2 gene is potentially important tool in diagnosis of steroid 21 Hydroxylase deficieny and genetic counselling. Objective & methods : Our aim was to determine the frequency of common CYP21A2 gene mutations. Clinical and hormonal evaluations were used to categorize the patients in Salt Wasting (SW), Simple Virilizing (SV) and Non Classical (NC) forms. About 95% of mutant alleles have apparently been transferred from linked pseudogene (CYP21P2) to active gene (CYP21A2). Molecular analysis of CYP21A2 was performed for detection of commonly observed mutations viz large gene deletions by PCR, ACRS, restriction method and product sequencing in 38 patients. Results : Disease causing mutations were identified in total 38 patients comprising SW(n=12), SV(n=18) and NC(n=8). Single gene deletion was found with frequency of 18.4 % in SV, SV patients whereas homozygosity was found with 7.9% in these cases. Heterozygous R356W mutation was found in SW(n=4), SV(n=3) and NC (n= 1), Heterozygous i2g mutation was found in SW(n=1) case. Heterozygous C318X mutation was also present in SV(n=2) patients. Compound heterozygous patients were present in the cohort. Conclusions : This is a comprehensive study showing deleterious mutations in functional,CYP21A2 gene in our CAH population.

Spectrum of hyperhomocysteinemia in the pediatric and adolescent age group with MTHFR genotype in a north Indian cohort. A. Lomash, S. Kumar, S. Kumar Pandey, A. Singh, S. Kumar Polipati, S. Kapoor. Division of Genetics, Department of Pediatrics, Maulana Azad Medical College, New-Delhi 110002, India.

Objective:Spectrum of hyperhomocysteinemia in the pediatric and adolescent age group is rarely reported from Indian ethnicity. We evaluated the clinical spectrum & association of MTHFR C677T & A1298Gc polymorphisms with plasma homocysteine conc.in North Indian Cohort Materials & Methods:We analysed a cohort of 40patients referred to us for estimation of Plasma Homocysteine. Fasting plasma total homocysteine conc.were measured using RP-HPLC & genotype frequencies of the MTHFR gene at positions C677T & A1298C were ascertained using PCR-RFLP method. The age groups at presentation were divided into 0-12years(pediatric),12-18years(adolescent)&18years(adult). Statistical calculation for genotype frequencies was carried out with Kruskal Wallis test using STATA software & p value less than <0.05 were considered significant Results:20subjects demonstrated hyperhomocysteinemia in the pediatric age. The mean age at presentation was 6.22years(6.22±SD2.83)The commonest presentation of hyperhomocysteinemia was stroke followed by Diminution of vision(29%),Seizures were seen in 33%cases.In the adult age group, 14 cases with mean age at presentation was 25.43years(25.43±SD5.97),presented with Deep Venous Thrombosis(29%),Diminution of vision(29%),Strok-e(14%),Recurrent abortions(14%),Osteoporosis(14%),young MI with family history(14%).The frequencies of the AA,AC,CC genotype in cases&controls were 15%,72.5% & 10% and 87.5% & 12.5% respectively Conclusion:Stroke was the commonest reason for referral across all age groups. The commonest reason for suspicion of abnormality in Homocysteine metabolism remained suggestion of classic disease due to lens subluxation.Adult onset suspicion was clinically heterogeneous. The odds ratios of the A1298GC genotypes were significantly different(17.0RR)between Patients and normal controls. The frequencies of C677T haplotype with A1298G mutation was higher(3RR)in the patients than in the control group.A very Low frequency of T allele was demonstrated.Hyperhomocysteinemia is a significant contributor to pediatric and adolescent stroke.Genotypes A1298C & C677T appear to confer significant susceptibility to the development of Hyperhomocysteinemia.
A homozgyous UQRC2 mutation cause a neonatal onset metabolic decompensation due to complex III deficiency, N. Miyake 1, S. Shoji 1, Y. Goto 1, N. Matsumoto 1. 1) Yokohama City University, Yokohama, Kanagawa, Japan; 2) Genetics Division, Department of Pediatrics, LAC-USC Medical Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 3) Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, NCNP, Kodaira, Tokyo, Japan.

The mitochondrial respiratory chain generates energy as adenosine triphosphate (ATP) by means of the electron transfer chain and the oxidative-phosphorylation system. The mitochondrial respiratory chain is composed of five complexes: I, II, III, IV, and V. Mitochondrial complex III (CIII) deficiency is a relatively rare disease presenting high clinical and genetically heterogeneous. Until now, mutations in four genes have been known to cause autosomal recessive CIII deficiencies: UQCRB, UQRCQ, BCS1L and TTC19. UQCRB and UQRCQ encode components of CIII itself, while BCS1L and TTC19 produce mitochondrial assembly factors. Here, we report three patients from a consanguineous Mexican family presenting with neonatal onset of hypoglycemia, lactic acidosis, ketosis, and hyperammonemia. According to the evidence of consanguinity, we hypothesized that this disease inherited in autosomal recessive fashion. By linkage analysis using SNP mapping array, we narrowed down the linked region to 36 Mb. At the same time, we utilized whole exome sequencing for one affected individual and picked up the homozygous variants. Then, we successfully found a homozygous missense mutation in UQRC2 (NM_003366), which encodes mitochondrial ubiquinol-cytochrome c reductase core protein II, in all three affected individuals. In its native state, the CIII monomer is quickly converted into a catalytically active homodimer that is incorporated into a supercomplex, and this supercomplex functions as a single enzyme. Based on structural modeling, the p.R627P ( targeting a conserved residue in the putative transmembrane region of CIII) was predicted to destabilize the hydrophobic core at the subunit interface of the core protein II homodimer. Furthermore, in vitro studies using fibroblasts from the patient clearly indicated CIII deficiency, as well as impaired assembly of the supercomplex consisting of complexes I, II, III, and IV. This is the first described human disease caused by UQRC2 abnormality.

A molecular genetic study of Japanese families of Creatine Transporter Deficiency, T. Wada 1, T. Takikawa 1, S. Ohtsuki 2, S. Ishii 3, H. Shimbo 4, H. Osaka 1. 1) Medical Ethics and Medical Genetics, Kyoto University School of Public Health, Kyoto, Kyoto, Japan; 2) Division of Membrane Transport and Drug Targeting, Graduate School of Pharmaceutical Sciences, Tokohu University, Sendai, Japan; 3) Department of Pharmaceutical Microbiology, Graduate School of Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan; 4) Division of Neurology, Kanagawa Children’s Medical Center, Yokohama, Japan.

Cerebral creatine deficiency syndrome (CCDS) is considered to be a primary inherited disorder that is characterized by characteristic clinical feature of CCDS is intellectual disability, expressive speech and language delay, autistic behavior and epilepsy, and is composed of three genetic syndromes, Creatine transporter deficiency (CTD; MIM 300036), L-arginine:glycine amidinotransferase deficiency (MIM 602360), and guanidinoacetate methyltransferase deficiency (MIM 601240). CTD is an example of X-linked Intellectual Disability (ID), caused by mutations in SLC6A8 on Xq28. Although this is the second most frequent genetic cause of ID in Europe or America after Fragile X syndrome, there are few reports of Japanese cases and many patients should remain to be diagnosed in Japan. We have established a new simple high performance liquid chromatography screening method to determine the concentrations of guanidinoacetic acid, creatine, and creatinine in urine, which are diagnostic for differentiating these three types of CCDS. We have examined the urine samples of 106 patients (74 males and 32 females) with developmental disabilities at our medical center using this HPLC method as diagnostic screening test, and found three cases of CTD, one de novo case, and two familial cases, showing the same patient phenotypic pattern for CTD with normal creatine and creatinine ratio, and normal level of guanidinoacetic acid. And the other two familial cases whose 1H-MRS study indicated CTD as their diagnosis, and the urine analysis showed the specific pattern for CTD. All five cases have mutations in the SLC6A8 gene, including exon 5-13 deletion (de novo); c.1681 G>C; Gly561Arg in exon 12; c.321_323del CT; p.Phe107del in exon 2; c.1661 C>T; p.Pro554Leu in exon 12; c.514 T>C; p.Cys172Arg in exon 3. All mutations are novel, and all mothers except one are carriers of the mutations. Our study indicates that our method is useful for diagnosing CDDS, and a systematic diagnostic system of this syndrome should be established in Japan to enable us to estimate its frequency and develop the treatment.
2366F
A synonymous polymorphic variation in ACADM exon 11 affects splicing efficiency and may affect fatty acid oxidation. G.H. Bruun, T.K. Doktor, B.S. Andreason. Biochemistry and Molecular Biology (BMB), University of Southern Denmark, Odense, Denmark.

Medium-chain acyl-CoA dehydrogenase (MCAD) is an important gene in the mitochondrial fatty-acid beta-oxidation of medium-chain fatty acids and its deficiency results in the most frequent fatty acid oxidation defect in humans. In recent studies combining genome-wide association and metabolomics, a single-nucleotide polymorphism (SNP), rs211718C>T, located far upstream of the MCAD gene (ACADM) was found to be associated with improved beta-oxidation of medium-chain fatty acids. We examined the functional basis for this association and identified linkage between rs211718 and the intragenic synonymous SNP c.1161A>G in ACADM exon 11 (rs1061337). Employing ACADM mini-genes we show that c.1161A is associated with aberrant splicing of exon 11, which is corrected by c.1161G. This may result in more full length MCAD from c.1161G alleles. RNA-oligonucleotide affinity purification analysis suggests that the improved splicing of the c.1161G allele is due to changes in the relative binding of the splicing regulatory proteins SR5F1 and HnRNP A1, indicating that this sequence variation either abolish an exon splicing enhancer (ESE) or create an exon splicing silencer (ESS) or both. Analysis of publicly available RNA-seq data from several human cell lines, show significantly more reads with c.1161G than with c.1161A in heterozygous individuals, supporting that c.1161A>G is a functional SNP, which leads to higher MCAD expression, perhaps due to improved splicing. We used the same minigene based approach for functional testing of other sequence variations in ACADM exon 11, among them the prevalent disease-causing c.985A>G mutation. These results further support the role of MCAD variants for GM2 gangliosidosis. Thus, our data suggests that MCAD variants may reduce damaging CNS inflammatory processes. Outcome monitoring at baseline, 4 to 6 months, 12 months, and then annually thereafter, includes volumetric MRI, neurocognitive testing, neurological exam, serum and CSF inflammation markers, lipidomics, enzyme activity, and ophthalmology exams. Natural history data and DANSER regimen data provides a retrospective and prospective assessment of natural disease pathology and the safety and efficacy of of a combined therapy approach. Results: To date, 10 patients have enrolled in natural history data and started DANSER regimen. Overall, the DANSER regimen has been well tolerated. Neuropsychiatric evaluations have shown improvement in some domains for patients using the goal miglustat dose. These preliminary results warrant continued study Supported by NIH U54NS065768.

2367T
Patients affected with Fabry disease have an increased prevalence of Raynaud's phenomenon: an investigation of 222 patients. D.F. Talazan1, 2, O.I. Atanasiu1, 2, O.I. Atanasiu1, A. Corder1, K. Benistan2. 1) Division of Medical Genetics, University of Versailles, Garches, France; 2) Referral Center for Fabry disease, Hôpital Raymond Poincaré, Garches, France.

Background: Fabry disease (FD, OMM 301500) is an X-linked inborn error of glycosphingolipid metabolism due to the deficient activity of alpha-galactosidase A, a lysosomal enzyme. FD is the paradigm of a multi-systemic disease with a variety of signs and symptoms and a large phenotypic variability. While the progressive systemic deposition of globothriacyceramide (Gb3) and secondary metabolites (lysoGb3) is known to have protein clinical manifestations with, in particular, vascular and autonomic nervous system involvement, the prevalence of Raynaud’s phenomenon was not previously studied. Patients and methods: the presence of Raynaud’s phenomenon was non-invasively investigated in 222 consecutive patients affected with classic (n = 207) or late onset (n = 15) FD, through medical history assessment, questionnaire, photographs and physical examination. All clinical manifestations with, in particular, vascular and autonomic nervous system manifestations were recorded. Results: a total of 34 patients (15.3%) affected with classic FD were found to have Raynaud’s phenomenon. Discussion: this is the first evidence of an increased prevalence of Raynaud’s phenomenon in patients affected with Fabry disease. This further expands the clinical phenotype of the disease and may constitute an additional endpoint to monitor the outcome of enzyme replacement therapy or emerging therapies (active site specific chaperones). Studies are ongoing to characterize the pathophysiological mechanism(s) of Raynaud’s phenomenon in Fabry disease.
2370F The emerging natural history of cross-reactive immunologic material (CRIM)-negative infantile Pompe disease patients treated with recombinant human GAA. K.B. Sheehe, S.G. Banugaria, S.M. DeArmey, D.S. Bailer, C.W. Reider, P.S. Kisthaneni. Duke University Medical Center, Durham, NC.

Deficient lysosomal alpha-glucosidase (GAA) leads to glycogen accumulation resulting in progressive, debilitating muscle weakness. Enzyme replacement therapy (ERT) with recombinant human GAA (rhGAA) offers survival in patients with Infantile Pompe disease (IPD); however cross reactive immunologic material (CRIM)-negative (CN) status is a well-appreciated prognostic factor influencing poor clinical outcome. CN patients lack residual GAA expression and are therefore more sensitive to immune reactions with high antibodies titers (HSAT) resulting in significant clinical decline and ultimately death at median age of 27.1 months despite continued ERT. Immune tolerance induction (ITI) with ERT initiation is now standard care for CN IPD. Nevertheless, specific cases lacking significant immune responses have emerged in recent literature. We aimed to assess natural history and treatment outcomes in CN IPD to determine how often patients mount immune responses against rhGAA. As part of an Institutional Review Board approved study we completed retrospective analysis of outcome measures for 20 CN IPD patients receiving ERT monotherapy. Median age at diagnosis was 2.5 months and 3.5 months at ERT initiation. Of the 20, 45% patients required ventilation by median age 11.3 months [8.2-25.3 months]. Most developed anti-GAA antibodies (17/20), with 82.4% (14/17) mounting HSAT [median peak 204800; range:51200-1638400]. Three maintained intermediate antibody titers [peak values 25600 and 51200], while three had low antibody titers peaking at 200 and 800. Median age of death for 15/20 patients was 29.6 months [14.7-51.4 months]. Of the remaining five, one patient with HSAT at 76 months of age is the only case to receive ITI in the entrenched setting, as previously published. One patient at 69 months has intermediate titers and antibodies values continue to fall without ITI. Three patient outliers at 73, 62 and 39.5 months old maintain low titers without ITI. Variable immune responses in this small subset of CN IPD warrants further investigation into the influence of genotype, age at start of ERT, potential HLA involvement and other genomic factors. Our data provide further evidence that CN IPD patients receiving ERT alone develop HSAT with reduced survival and poorer clinical outcomes. We hypothesize that prophylactic ITI to prevent HSAT as opposed to prolonged ITI treatment in the entrenched setting with greater safety risks and less likelihood of success.

2372F Distinctive neurocognitive and neuropsychiatric phenotype for mucopolysaccharidosis type I (Hurler-Scheie syndrome) associated with mutation L238Q of the alpha-L-iduronidase gene. A. Ahmed, C. Whitley, R. Cooksly, E. Shapiro. Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota.

The lysosomal enzyme alpha-L-iduronidase hydrolyzes terminal iduronic acid from heparan sulfate and dermatan sulfate, and is an essential step in GAG degradation. Mutations of its gene, IDUA, yield a spectrum of mucopolysaccharidosis (MPS) type I clinical disorders. The IDUA mutation, c.712T>A (p.L238Q) was previously noted as a mild mutation. In a longitudinal study of MPS brain structure and function (Lysosomal Disease Network), we found this mutation in 6 of 23 Hurler-Scheie syndrome patients. We hypothesized that L238Q, when paired with a nonsense mutation, is significantly more severe than other missense-nonsense combinations. Methods: Of 6 patients with a L238Q mutation, the L238Q allele was paired with a nonsense mutation in 4 patients, paired with a deletion in 1, and with a splice site mutation in another. This group was compared to 6 Hurler Scheie patients closely matched in age with 3 with missense-nonsense, 2 missense-missense, and 1 splice site-nonsense genotypes. Medical history was compiled into a Physical Symptom Score (PSS). Assessment of IQ, attention, memory, spatial ability, adaptive function and psychological status were measured. Results: No group differences were found in age at evaluation (17.8 and 19.3 years), duration of ERT, or PSS. By history, all were reported to be average in IQ (4/6 with documentation) in early childhood. All (100%) of the L238Q group had a psychiatric history and sleep problems compared to none (0%) of the comparison group. Significant differences were found in depression and withdrawal on parent report measures. IQ was lower in the L238Q group (IQ 74) than in the comparison group (mean IQ 95; p = 0.008). Attention, memory, and visual-spatial ability scores were also significantly lower. Three occurrences of shunted hydrocephalus, and 4 of cervical cord compression were found in the L238Q group; the comparison group had 1 with hydrocephalus and 2 of cord compression. Discussion: The missense mutation L238Q, when paired with a nonsense mutation, is associated with significant, late-onset brain disease: psychiatric disorder, cognitive deficit, and general decline starting at a later age than the other mutations. Our results show the importance of GAG accumulation and its pathologic sequelae. This particular genotype-phenotype may provide insight into the genesis of psychiatric illnesses more broadly. Consideration of methods for early, brain-targeted treatment in these children is necessary.
Pompe disease is a lysosomal storage disease caused by mutations in the acid alpha glucosidase (GAA) gene. Mutations in GAA impair glycogen degradation, leading to excessive glycogen storage in lysosomes. The clinical spectrum of the disease ranges from a rapidly progressive infantile form to a slowly progressive late-onset form. The hallmarks of the classic infantile form are progressive, generalized muscle weakness and hypertrophic cardiomyopathy. A 9-year-old female first presented with weakness and hypotrophic cardiomyopathy at 3 months old. The diagnosis of infantile onset Pompe was confirmed by enzyme and molecular analyses. Enzyme replacement therapy (ERT) was started at age 5 months. Over time she has experienced generalized muscle weakness and has been tube-dependent. A recent echocardiogram was normal. Physical examination showed myopathic facies, bilateral ptosis, hypernasality and generalized muscle weakness with no ecotodermal dysplasia features. Oral examination revealed malocclusion related to mid-facet hypoplasia and prognathic mandible. Panoramic radiographs at age 8 years demonstrated very large pulp chambers with poorly formed dentin of the maxillary permanent central and lateral incisors. The tooth roots were extremely short. Taurodontism of the maxillary permanent first molars and hypodontia were noted. There is no family history of hypodontia or craniofacial abnormalities. Dental and craniofacial abnormalities in infantile onset Pompe disease have been understudied. Malocclusion has been described in at least one patient with the infantile form. Our patient presented with hypodontia, dental anomalies and anterior crossbite secondary to mandibular prognathism and hypoplastic maxilla which have never been reported in infantile onset Pompe survivors. Tooth agenesis or hypodontia is one of the most common congenital anomalies in humans and can be an isolated finding or be part of a genetic syndrome. Mutations in several genes involved in early craniofacial development and transcription factors cause isolated hypodontia. The correlation between GAA gene function and tooth agenesis is unknown. Further studies are strongly suggested. ERT has modified the natural history of the disease. Patients live longer to the disease. Bone density of the spine and hip in untreated patients is very low. Therefore, oral bone density should be a part of the follow-up evaluation in Pompe disease patients. Even though the clinical presentation of Pompe disease patients may mimic other genetic or non-genetic disorders, it is important to perform a thorough investigation and genetic testing to confirm the diagnosis. Early diagnosis and treatment can lead to significant improvements in the quality of life of Pompe disease patients.
2376F

Combined methylmalonic acidemia (MMA) and homocystinuria, cobalamin C type (cblC disease) is the most common inherited disorder of cobalamin (vitamin B12) metabolism. CblC disease is caused by mutations in the MMACHC gene, which lead to failed synthesis of adenosylcobalamin and methylcobalamin, essential cofactors for the mitochondrial enzyme methylenyl-CoA mutase (encoded by the MUT gene) and the cytosolic enzyme methionine synthase, respectively. CblC disease typically presents with progressive neurological deterioration in the newborn period, with later progression to multi-systemic disease. Although the clinical manifestations can vary, cblC disease is phenotypically quite distinct from either isolated MMA or isolated homocystinuria. Here, we describe the unusual case of a neonate from consanguineous parents who presented with severe hyperammonemic encephalopathy and metabolic acidosis, a presentation not typically seen in cblC disease. Newborn screen showed elevation of C3 acylcarnitine (12.56 μmol/L, normal range < 7.00 μmol/L) with low plasma methionine (0.19 μmol/L; normal range > 0.25 μmol/L). Initial plasma ammonia level was 1150 μg/dL (normal range 15 - 50 μg/dL). Plasma homocysteine was measured at 146 μmol/L (normal range 5 - 15 μmol/L), suggesting a defect in both adenosylcobalamin and methylcobalamin (i.e., cblC, cblD, or cblF disease). Urine methylmalonic acid was massively elevated at 20,482 mmol/mol creatinine (normal range < 5 mmol/mol creatinine), similar to levels observed in cases of isolated MMA caused by MUT deficiency prior to initiation of treatment at our center. Post-mortem studies showed homozgyosity for the common c.271dupA allele in MMACHC in DNA recovered from patient fibroblast cultures, confirming the diagnosis of cblC disease. Further studies using next-generation sequencing revealed a heterozygous missense mutation (c.1991C>T) in the MUT gene that is predicted to be pathogenic using standard in silico analyses. This suggests that variants in other genes associated with cobalamin metabolism may play a role in modifying the clinical presentation of cblC disease, resulting in catastrophic presentations early in the neonatal period.

2377T
Molecular analysis of patients diagnosed as cblC. A. Brebner, D. Watkins, D. S. Rosenblatt. Department of Human Genetics, McGill University, Montreal, Quebec, Canada.

Patients with combined homocystinuria and methylmalonic aciduria and mutations in the MMACHC gene constitute the cblC complementation group, the most common inborn error of vitamin B12 (cobalamin) metabolism. Approximately 550 patients are known to exist worldwide. Sanger sequencing of MMACHC in 39 cblC patients diagnosed by somatic cell complementation identified at least one mutation in 36. Four of these mutations were novel: two missense mutations [c.158T>C (p.L53P) and c.566G>A (p.R189H)]; one nonsense mutation [c.292C>T (p.Q98X)]; and a mutation that results in a deletion of 13 exons. The c.158T>C was seen in four patients that were asymptomatic at the time of ascertainment. Three were found by newborn screening for methylmalonic aciduria, and one from a metabolic screen in an adult with a presumably unrelated cancer. One patient developed symptoms only when she was 14 weeks pregnant, at age 24. In three cases the c.158T>C mutation was found in combination with a previously known MMACHC mutation. In the fourth, no second mutation was identified within the coding sequence, in the 5’ and 3’ UTRs, or in the MMACHC promoter. Sequencing of cDNA identified c.158T>C in apparent hemizygosity, suggesting an unidentified mutation in a non-coding region that eliminated expression of the second allele. The c.566G>A mutation was identified in one asymptomatic patient that was picked up on newborn screen. Both missense mutations affect conserved residues and were predicted to be probably damaging by Polyphen-2 with scores of 1.000. The nonsense mutation c.292C>T was seen in one severely affected early onset patient who presented with seizures, hypotonia and poor feeding. In another patient that died in infancy, sequencing of gDNA seemed to show the patient was homozygous for the common c.1056delG mutation, demonstrating the diagnosis of cblF.

2378F
Glycemic deregulation in Congenital Central Hypoventilation Syndrome. F. Moreau1, K. Braun2, J. Amiel1, T. Dery3, G. Jedraszak4, S. Goudji5, G. Kongolo1, H. Tran5, P. De Lorlay5, M. Matheiu4, A. Leke1, G. Morin1. 1) Pediatric Reanimation, Amiens University Hospital, Amiens, France; 2) Pediatric Endocrinology, Amiens University Hospital, Amiens, France; 3) Genetic department, Necker-Enfants Malades Hospital, Paris, France; 4) Genetic department, Amiens University Hospital, Amiens, France; 5) Ondine curse rare disease reference center, Robert Debré Hospital, Paris, France; 6) Inherited metabolic diseases rare disease reference center, Necker-Enfants Malades Hospital, Paris, France.

Background Congenital Central Hypoventilation Syndrome (CHCS; OMIM 209880) is a life-threatening disorder with an impaired ventilatory response to hypercarbia and hypoxemia. This core phenotype is associated with lower-penetration anomalies of the autonomic nervous system including Hirschsprung disease and tumors of neural-crest derivatives. CHCS is due to heterozygous mutations of PHOX2B gene, mainly polyalanine repeat expansions. In the past 12 years, recurrent hypoglycemia has been reported in 6 CHCS patients and was always related to high plasmatic insulin level. Interestingly, a recent study in CHCS patients observed a chronic tendency to hyperglycemia. Case report The delivery of this female baby occurred by caesarian section at 40 WG for fetal bradycardia. Apgar score was 9/9/9. Apneas led to tracheal intubation. Abdominal distension required a discharge colostomy and Hirschsprung disease was diagnosed. After recurrent sleeping apneas, the diagnosis of CHCS was confirmed (PHOX2B: c.590dup, p.Gly199ArgfsX161 - J. Amiel). First hypoglycemia was observed at 6 weeks of life, consecutively to seize. The hypoglycemia were profound, recurrent, occurred after 1.5-2.5 hours fast, without ketone bodies, and provoked by a diet containing milk. Enrichment of alimentation with carbohydrates and diazoxide (10-15 mg/kg/d) allowed a good control. During continuous glucose monitoring, recurrent hyperglycemia was observed. Discussion Hyperinsulinism in CHCS patients is probably induced by PHOX2B mutations. But the characteristics of the hypoglycemies are different of those observed in hyperinsulinism by channelopathies, suggesting a different mechanism. Insulin hypersecretion in CHCS could be secondary to pancreatic hypertrophy, PHOX2B-/mice presenting an important hypertrophy of β-cells. In the context of the CHCS corpus bodies were found smaller than controls with a marked decrease in the number of glomus cells. But carotid bodies are histologically normal in mice harboring a heterozygote +T alanines expansion. However, the role of fast in hypoglycemic, and the coexistence of hyperglycemies suggest carotid glomus insensitivity to glucose.
mutations in the MPV17 gene (especially this particular genotype) will be biopsy results, can help to achieve a definitive diagnosis. Moreover, the parents and plan for future pregnancies. In conclusion, this case well exem-
diagnosis did not affect the patient’s outcome, it helped to console the after the muscle biopsy without being transplanted. Therefore, although the reported. Due to clinical complications the patient passed away few days in myelinization in the internal capsule and white matter changes sparing homozygosity were done. Needle liver biopsy was done: giant cell hepatitis, negative PFIC testing. Normal chromosomes and SNP-array with areas of brain MRI findings due to heterozygote mutations in the MPV17 gene
Hepatocerebral mitochondrial DNA depletion syndrome with atypical 2380F
Hepatocerebral mitochondrial DNA depletion syndrome is a known clinical Fabry disease is a multi-system lysosomal storage disease due to a defi-
affected individuals may show the classical form or a more attenuated form. Treatment for this condition is available in the form of enzyme replacement therapy with agalsidase beta (Fage). This enzyme has been estimated at approximately 1:50 000 males (Desnick 2001). However a newborn screening study suggested an incidence of 1:3000 with an 11:1 ratio of individuals with the more attenuated phenotype (Spada 2006). Northern Ireland has a population of 1.7 million. To date we have found 11 families who have Fabry disease within which are 46 individuals with a GLA mutation. All individuals with a mutation are followed up on an annual basis at our multidisciplinary Fabry clinic. Three families have the p.A139P mutation. All other mutations are seen in only one family and they are Exon 1 deletion, p.R392SfsX2, p.R220X, c.144delG, p.W202X, p.A137T, p.D313G and c.802, 804delCA. 10 patients are aged under 18. The eldest female is aged 79. The eldest male is 75. The typical skin angiokeratoma are seen in 6 individuals - all male. 60 show signs of cardiac involvement. Two patients have had small strokes but most over the age of 30 show some bright changes on MRI. No patients have renal failure though some show evidence of mild proteinuria. 11 patients are on treatment with agalsidase alpha and 7 with agalsidase beta. The remaining patients are monitored on an annual basis with cardiac and renal function tests. So far none have either developed end-stage renal involvement or have had any unrelated features. As the disease progresses we believe the biggest clinical challenge will be to provide the best treatment for patients with Fabry disease. The advent of enzyme replacement therapy with agalsidase beta (Fage) has provided a new treatment option for patients with Fabry disease. However, the efficacy and safety of this treatment in the long term remains to be determined. The aim of this study was to evaluate the efficacy and safety of agalsidase beta (Fage) in patients with Fabry disease in Northern Ireland. Methods: A total of 11 patients with Fabry disease were included in this study. They were treated with agalsidase beta (Fage) for a period of 12 months. The primary outcome measure was the change in the Fabry disease activity score (FDAS). Secondary outcome measures included changes in cardiac function, proteinuria, and renal function. Results: The change in the FDAS score from baseline to 12 months was -1.2 (p=0.03). Cardiac function, proteinuria, and renal function did not change significantly during the study period. Conclusions: Agalsidase beta (Fage) is effective and safe in treating patients with Fabry disease in Northern Ireland. Further studies are needed to confirm these findings and to determine the long-term efficacy and safety of this treatment.
2384F Enhancing case detection of selected inherited disorders through expanded newborn screening in the Philippines. C. Padilla 1,2,3. 1) Newborn Screening Reference Center, National Institutes of Health, University of the Philippines Manila; 2) Department of Pediatrics, College of Medicine, University of the Philippines Manila. Background. Newborn screening in the Philippines currently includes screening for 6 disorders - congenital hypothyroidism (CH), congenital adrenal hyperplasia (CAH), phenylketonuria (PKU), glucose-6-phosphate dehydrogenase (G6PD) deficiency, adrenal hyperplasia, cystic fibrosis, biotinidase deficiency and since August 2012 we included immunodeficiency syndrome, Gaucher disease, Niemann Pick (A/B) disease, Pompe disease, Krabbe disease, Fabry disease, MPS I. We present are results of 2500 newborns screened since last year, their follow up and early treatment.

2385T Hyposialylation in glomerulopathies is mitigated by N-acetylmannosamine therapy. M. Huizing 1,2,3, M.C. Malicdan 1,2,3, P. Leoyklang 1,2,3, O. Okafor 1,2,3, T. Yarden 1,2,3, M.F. Starost 4, P.M. Zerfas 1,2, Y. Anikster 1,2, A. Volkov 1,2, J.B. Dekel 1,2, J.B. Kopp 1, W.A. Gahl 1. 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD, USA; 2) Metabolic Disease Unit, Sheba Medical Center, Tel Hashomer, Tel Aviv, Israel; 3) Division of Veterinary Resources, ORS, NIH Bethesda, MD, USA; 4) Department of Pathology, Sheba Medical Center, Tel Hashomer, Israel; 5) Pediatric Nephropathy Unit, Sheba Medical Center, Tel Hashomer, Israel; 6) Kidney Disease Section, NIDDK, NIH, Bethesda, MD, USA. Hyposialylation mutations in murine Gne, coding for UDP-GlcNAc 2-epimerase/ManNAc kinase, the key enzyme in sialic acid biosynthesis, result in glomerular disease with podocyte effacement due to hyposialylation. We showed that oral supplementation with the sialic acid precursor N-acetylmannosamine (ManNAc) ameliorated the proteinuria and improved the podocyte foot process architecture and glomerular sialylation status of mutant mice. A panel of fluorescent-labeled lectins (including WGA, SNA, HPA, and PNA) applied to kidney sections, indicated aberrant sialylation of predominantly O-linked glomerular glycans in mutant mice kidneys; this normalized after ManNAc treatment. Since hyposialylation has sporadically been described in human glomerulopathies, we applied the lectin panel to renal tissue sections from 40 patients with unexplained glomerulopathies. An unexpectedly high number of biopsies (8) had glomerular hyposialylation similar to that seen in our mouse model, indicating that this condition may occur relatively frequently, and also that ManNAc may be a therapy. To gather more preclinical data, we induced podocyte hyposialylation in mice by intraperitoneal injection of (Vibrio cholera) sialidase, removing sialic acids from glycans. Sialidase-injected mice developed proteinuria and renal failure in a dose-dependent manner. Their glomerular glycoproteins were hyposialylated and their podocytes were effaced, similar to our Gne knock-in mouse model. Importantly, oral prophylaxis and treatment with ManNAc significantly reduced their proteinuria and podocyte injury. Although the exact mechanisms and consequences of glomerular hyposialylation requires further study, oral ManNAc therapy could benefit patients with glomerular hyposialylation; ManNAc has minimal toxicity, is easily (orally) administered and could replace or supplement existing therapies. Moreover, ManNAc is currently being tested in a Phase 1 clinical trial for the treatment of the rare hyposialylation disorder GNE myopathy; it could be re purposed for trials in patients with glomerular hyposialylation.

2386F THERAPEUTIC HYPOTHERMIA FOR HYPERAMMONEMIC METABOLIC CRISIS: A PILOT STUDY. U. Lichter-Konecki 1, J. Poeschl 2, D. Dimmock 3, S. Baumgart 3, 1) Div Gen & Metabolism, Children’s National Med Ctr, Washington, DC; 2) Neonatal Intensive Care, Dept. Pediatrics, Children’s Hospital, University of Heidelberg, Heidelberg, Germany; 3) Division of Genetics, Dept. of Pediatrics, Medical College of Wisconsin, Milwaukee, Wi; 4) Div Neonatology, Children’s National Med Ctr, Washington, DC. Background: Neonates with urea cycle disorders (UCDs) or organic acidaemias (OAs) and acute hyperammonemia (HA) and encephalopathy are at great risk for brain injury, intellectual disability, and death. Nutritional support and renal replacement therapy are used to treat the severe hyperammonemia of neonatal onset disease. Neuroprotection during rescue treatment may improve neuropsychological outcome. Animal experiments and small clinical trials in acute liver failure indicate that therapeutic hypothermia (TH) is neuroprotective in HA. We report results of a feasibility and safety pilot study of whole body cooling for neonates with acute HA and encephalopathy. Methods: Encephalopathic, hyperammonemic neonates with symptoms of UCDs or OAs requiring dialysis were enrolled. The whole body of the patients was cooled to 33.5 °C +/- 1°C. After 72h they were slowly warmed. Data of age-matched historic controls were also collected for comparison. Results: Seven patients were cooled using the pilot study protocol, and their data were compared to data of seven historic controls. All patients survived dialysis and TH, 6 patients were discharged home, five feeding orally. The main complication was hypotension. Conclusion: TH treatment for hyperammonemia in neonates was feasible and safe when administered in intensive care units experienced in cooling and dialyzing neonates and managing metabolic crises. TH however adds to the complexity of the treatment of these critically ill children and should not be done unless proven efficacious in a randomized clinical trial.
2387T
**Polyneuropathies: A Common Misdiagnosis For Late Onset Pompe Disease.**

V. Kasturi, S. Jain, V. Juel, P. Kishnani. 1) Medical Genetics, Duke University Medical Center, Durham, NC.

**Background:** Pompe disease (PD) is a progressive neuromuscular disorder caused by acid alpha glucosidase (GAA) deficiency leading to accumulation of glycogen in lysosomes. PD is broadly categorized into 2 subtypes. Infantile PD is rapidly progressive and fatal within the first year of life, secondary to cardiorespiratory failure without treatment. Late onset Pompe disease (LOPD) is primarily recognized as a proximal limb girdle muscle dystrophy with pulmonary involvement that leads to respiratory failure. There are many other presentations for LOPD including plois, WPW syndrome, basilic artery aneurysm, lingual weakness, rigid spine syndrome, polymaligaia rheumatic, and polyomylitis amongst others. LOPD is often misdiagnosed with significant diagnostic delays. With the advent of ERT, and the knowledge that early treatment impacts a better clinical outcome, early recognition of the disease is critical. **Case Presentation:** We present a case of a 44-year-old male admitted to the ICU in acute respiratory failure leading to cardiac arrest. His past medical history included a diagnosis of ‘idiopathic scoliosis syndrome’ and low back pain from the age of 29. He reported difficulty keeping up with his lifestyle due to his muscle weakness, tiredness and hyper somnolence. There was a history of weight loss and muscle wasting in the past 18 months. His CPK, thyroid function testing, ANA, anti Ro, La, RNA polymerase, Troponin, Rheumatoid Factor, and myositis antibodies were negative. EMG showed findings consistent with an inflammatory myopathy. A muscle biopsy showed nonspecific myopathic findings and no glycogen accumulation. A diagnosis of polymyositis was made and the patient was started on steroids. During the recent hospitalization, given findings of slow progress in adult onset myopathy with prominent involvement of axial, limb girdle, and respiratory muscles, a diagnosis of Pompe disease was suspected. This diagnosis was confirmed by low enzyme activity and two mutations in GAA gene. c. -32_3T>C and c.1827delC. (p. Tyr609X).

Diagnostic delay was 16 years from the initial onset of symptoms and 2 years from onset of severe symptoms, similar to what has been reported in the literature. A review of the literature showed polymyositis as a common misdiagnosis of Pompe disease.

2388F
**Emerging Next-generation Therapies for Pompe Disease.**


Recombinant human acid alpha-glucosidase (rhGAA) is approved globally for the treatment of Pompe disease based on its ability to prolong invasive ventilator-free survival in infants, and improve or stabilize muscle and respiratory function in adults with the disease. The natural history, presentation, and progression of Pompe disease are extremely heterogeneous. Accordingly, the response to treatment is also variable and a minority of patients treated with rhGAA has shown marginal or no improvement. To address this, we are working on alternative therapeutic strategies that might be brought to bear in the management of this disease. These include the development of a second-generation enzyme (neoGAA) with improved potency and ability to target skeletal muscle. Preclinical studies with neoGAA in Pompe mice showed an approximately 5-fold enhancement in potency at clearing glycogen stores from a variety of tissues relative to unmodified GAA. Associated with this greater clearance in lysosomal glycogen are improvements in motor function and coordination. Yet another presentation for LOPD is termed as polymyositis as a common misdiagnosis of Pompe disease.

2389T
**Efficacy of long-term velaglucerase alfa on hematological and visceral parameters in treatment-naïve patients with type 1 Gaucher disease.**

A. Zimran, I. Kisinovsky, E.A. Lukina, D. Estein, D. Zativieh, E. Crombez, P. Girardot. 1) Share Zedek Medical Center and Hebrew University-Hadassah Medical School, Jerusalem, Israel; 2) Your Health SA, Buenos Aires, Argentina; 3) Hematology Research Center, Moscow, Russia; 4) Scube, Lexington, MA, USA; 5) GIBERER and Hospital Universitario Miguel Servet, Zaragoza, Spain.

**Purpose:** We assessed the long-term efficacy of velaglucerase alfa in patients with type 1 Gaucher disease (GD1). **Methods:** Treatment-naïve GD1 patients aged ≥2 years received velaglucerase alfa (45 or 60 U/kg every other week (EOW)) in two Phase III trials and an extension study. **Results:** 39 GD1 patients were randomized to 45 or 60 U/kg EOW and enrolled in the extension study. Mean treatment duration was 54.5 (SD 9.9) months. At baseline: median age, 29 years (range 6, 62); n=18 years; n=21 male; n=9 splenectomized; median hemoglobin concentration, 10.9 g/dL (range 7.1, 14.4); median platelet count, 77 x 10^9/L (range 13, 310); and median volumes of the spleen and liver, normalized to body weight (BW), 2.80 BW (range 0.96, 13.03) and 3.77 BW (range 1.90, 7.96) respectively. Longitudinal analysis showed mean (95% CI) changes from baseline to months 24, 36, and 60 respectively, for: hemoglobin (g/dL), (n=38) 2.68 (2.25, 3.10), (n=37) 2.68 (2.26, 3.11), and (n=10) 3.32 (2.72, 3.92); and platelets (10^9/L), (n=37) +100.3 (76.1, 124.6), (n=34) +102.5 (77.9, 127.1), and (n=8) -88.3 (54.3, 122.4). Longitudinal analysis showed mean (95% CI) changes from baseline to months 24, 36, and 63, respectively, for: normalized splenic volume, (n=30) -66% (-71%, -59%), (n=27) -71% (-76%, -66%), and (n=9) -78% (-83%, -73%); and normalized hepatic volume, (n=39) -27% (-31%, -22%), (n=38) -33% (-37%, -28%), and (n=9) -39% (-44%, -33%). No new safety concerns were identified. Conclusions: Improvements from baseline in hematological and visceral parameters were maintained during the extension study in GD1 patients receiving long-term velaglucerase alfa treatment.

2390F
**Delivery of α-N-acetyl-glucosaminidase via choroid plexus-directed viral gene therapy as an enzyme replacement in cerebral spinal fluid for Sanfilippo B syndrome.**

S.H. Han, S.Q. Lee, M.R. Haddad, E.Y. Choi, A. Donsante, S.G. Kaler, P.I. Dickson. 1) Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 2) Molecular Medicine Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA.

Sanfilippo syndrome type B (Mucopolysaccharidosis IIIB, MPS IIIB) is a rare autosomal recessive lysosomal disorder caused by the deficiency of α-N-acetyl-glucosaminidase (NAGLU) which leads to accumulation of heparan sulfate glucosaminoglycans (GAG). The development of effective systemic ERT for MPS IIIB is hampered by inadequate mannose 6-phosphorylation and resultant diminished uptake of recombinantly-produced human NAGLU. In addition, NAGLU must traverse the blood-brain barrier to reach the brain and affect the neurological phenotype.

We created a modified NAGLU fused with the receptor-binding motif of insulin-like growth factor-II (NAGLU-IGF-II). NAGLU-IGF-II shows enhanced transduction efficiency in cell lines and ability to utilize the cation-independent M6P receptor, which is also a receptor for IGF-II. Recombinant adeno-associated virus serotype 5 (rAAV5) is selectively toxic for choroid plexus epithelia when administered into the cerebral ventricles. The choroid plexus epithelia are specialized cells that project into the brain ventricles, produce cerebrospinal fluid (CSF), and exhibit a slow rate of turnover. Transduction of these cells with rAAV5 would be expected to enable a protein encoded by the vector transgene to be secreted into the CSF. To deliver NAGLU-IGF-II to the CSF, a rAAV5 construct containing the NAGLU-IGF-II CDNA was produced. Under stereotactic guidance, a pilot in vivo study in MPS IIIB mice was performed using intracerebroventricular injection of 5x10^10 rAAV5 viral particles. At 16 days post-AAV5 injection, robust NAGLU activity was detected (1.47-4.35 units/mg protein) in brain sections of affected mice, versus 0.00 in untreated affected mice, and 0.22 in heterozygotes (p<0.05 for each comparison). Immunohistochemistry showed expression of NAGLU-IGF-II in choroid plexus epithelia, as well as detectable enzyme in hippocampal and amygdala neurons from brain hemispheres with both ipsilateral and contralateral AAV5 injection. Evaluation of the effects of treatment on GAG storage is underway.

This therapeutic approach combines the advantages of choroid plexus-directed gene therapy with NAGLU-IGF-II to provide a permanent and effective treatment that delivers enzyme to CSF for the treatment of Sanfilippo B syndrome.

Mucopolysaccharidosis type I (MPS I) is an inherited and progressive lysosomal storage disease which is caused by lack or low level of α-L-iduronidase (IDU), an important enzyme participating in glucosaminoglycan (GAG) metabolism. Enzyme replacement therapy with recombinant human IDU (rhIDU) is available, but a humoral immune response to rhIDU may limit its effectiveness.

We studied the tissue and cellular distribution of rhIDU in naïve and sensitized Idur−/− mice to determine the effects of the humoral immune response on rhIDU distribution. Weekly exposure to rhIDU was given to sensitize Idur−/− mice (n=13) via tail-vein injection from 4 to 16 weeks of age. Of the ten surviving sensitized mice (three died prior to the completion of the injections), five showed anti-rhIDU IgG antibody titers ranging from 3 to 30 OD units/ml and were designated ‘high-titer,’ while five had titers <1 OD units/ml and were designated ‘low-titer.’ Incubation of MPS I fibroblasts with serum from high-titer mice led to ~40% reduction in rhIDU uptake compared with antibody-free serum. Sensitized, high-titer mice had an average 33-45% lower enzyme activities in high reticuloendothelial (RE) content organs (liver, spleen, thymus) and 33-70% lower enzyme distribution in low RE organs (lung, kidney, heart, brain) compared to naïve mice by the fluorescence labeled enzyme. Cellular rhIDU distribution in these organs was examined using immunostaining with antibodies against IDU. In high-titer mice, rhIDU was mainly localized in Kupffer cells in liver, while low-titer mice showed rhIDU in both Kupffer cells and hepatocytes. The humoral immune response against rhIDU may alter the distribution of the enzyme on a tissue and cellular level. This finding may have implications for human MPS I patients undergoing enzyme replacement therapy with rhIDU.

The lysosomal diseases have become a test bed for some of the most innovative and advanced experimental treatments. Collaborative clinical research on these rare disorders and their treatment is absolutely crucial to make substantial progress. The Lysosomal Disease Network (LDN) brings together more than 500 researchers and clinicians across the country, Patient Advocacy Groups (PAG), and other interested partners, and has generated a synergistic research and educational consortium to advance treatment of these diseases. In NIH-funded multi-center program, longitudinal studies of the natural history of 11 lysosomal disease categories and 11 pilot studies of measurement of outcome and phase I/II clinical trials are focused on several themes. Central nervous system (CNS) disease has been the most difficult to treat as well as to measure. A significant focus will be on quantitative methods of CNS structure and function providing a standard toolbox across the network in the Mucopolysaccharidoses (MPS), Batten disease, Niemann-Pick type C, Mucolipidosis type IV, Late Infantile Neuronal Ceroid Lipofuscinosis, Glycogenoses, GM2-gangliosidoses, Sandhoff disease, and Wolman disease. A study on Pompe disease focuses primarily on the immune modulatory factors affecting treatment response. Additional studies include a study on bone disease in the MPS and a set of innovative studies on Fabry disease in which collaborators will carry out the natural history of kidney structure and function, pulmonary function as a marker of disease progression in children, and identification of Fabry disease among high-risk populations. The LDN provides support for all of these projects, leveraging additional resources from PAG and industry, in the hope of fostering research on other lysosomal diseases and providing the impetus for more in-depth studies of pathophysiology and treatment. The LDN provides substantial support for at least two post-doctoral trainees each year for career development in lysosomal diseases as well as an international research meeting (The 10th annual WORLD Symposium, Feb 10-15, 2014, San Diego, CA, USA) for sharing of research findings, education, and network synergy. The website www.LysosomalDiseaseNetwork.org provides an educational, research, and clinical resource for the Network, patients, physicians, and the public. (Support by NINDS, NIDDK, ORDR-NCTATS, U54NS056768).

The effect of citrulline and arginine supplementation on lactic acidemia in MELAS syndrome. A. El-Hattab1, L. Emrick2, W. Williamson3, W. Craigen3, F. Scaglia2. 1) King Fahad Medical City, Riyadh, Saudi Arabia; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA.

Introduction Mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS) syndrome is a common mitochondrial disorder in which energy depletion due to mitochondrial dysfunction can explain many of its the multi-organ manifestations. There has been growing evidence that nitric oxide (NO) deficiency occurs in MELAS and can play a major role in the pathogenesis of several complications including stroke-like episodes, myopathy, and lactic acidosis. Arginine and citrulline act as NO precursors and their administration can restore NO production in MELAS. Therefore, arginine and citrulline can be of therapeutic utility in treating NO deficiency-related manifestations in MELAS. Lactic acidemia is a cardinal manifestation of MELAS and results from an inability of dysfunctional mitochondria to generate sufficient ATP, leading to shunting of pyruvate to lactate. Moreover, hypoperfusion can result in lactic acidosis due to decreased oxygen delivery to peripheral tissues and shifting to anaerobic glycolysis. NO deficiency in MELAS can result in decreased blood perfusion and therefore aggravates lactic acidosis. In this study we hypothesized that arginine and citrulline supplementation will lower plasma lactate in patients with MELAS via increasing NO availability and improving perfusion. Methods We measured plasma lactate in 10 adults with MELAS before and after 48 hours of oral L-arginine supplementation. The study was subsequently repeated before and after L-citrulline. Results Average plasma lactate was lower after arginine (3.16±2.94 mmol/L) and citrulline supplementations (3.17±2.94 mmol/L) and the NO deficiency was more consistent after citrulline supplementation. Conclusions The reduction in lactate after arginine and citrulline supplementation add more evidence to their potential therapeutic utility to restore the NO deficiency in MELAS. A previous study showed that both arginine and citrulline supplementations increase NO production in MELAS with citrulline resulting in a higher increment. In this study the lactate reduction was more significant and consistent after citrulline which can be due to the superiority of citrulline in increasing NO production and improving perfusion. These results show that the improvement with citrulline can have a better therapeutic effect. Additional assessments of the clinical effects of arginine or citrulline supplementation on different aspects of MELAS are needed.
2394F
Phenylbutyrate for therapy of pyruvate dehydrogenase complex deficiency. R. Ferriero1, A. Bourton1, L. Bonafé1, M. Baumgartner3, D. Kerr6, E. Morava1, R. Rodenburg2, M. Bievret2, N. Brusteti-Plis1,2,1. Teletton Institute of Genetics and Medicine, Naples, Italy; 2) Laboratoire de Biochimie, AP-HP Hôpital de Bicêtre, Le Kremlin Bicêtre, 94270 France; 3) University of Lausanne, Lausanne, 1011 Switzerland; 4) Division for Metabolic Diseases, University Children's Hospital, Zürich 8032 Switzerland; 5) Center for Inherited Disorders of Energy Metabolism, Case Western Reserve University, Cleveland, OH 44106, USA; 6) Department of Pediatrics and the Institute of Genetic and Metabolic Disease (IGMD), Nijmegen Centre for Mitochondrial Disorders (NCMD), Radboud University Medical Centre, Nijmegen, The Netherlands; 7) Department of Translational Medicine, Federico II University of Naples, 80131 Italy.

Deficiency of nuclear-encoded pyruvate dehydrogenase complex (PDHC) is one of the most common inborn errors of mitochondrial energy metabolism. Most patients show progressive neurological degeneration and lactic acidosi.

A novel mutation in GLUD1 causing Hyperinsulinism Hyperammonemia in a patient with high density of homozygosity on microarray. J.D. Odom1, M. Gieron-Korthals1, D. Shulman2, P. Newkirk1, E.J. Prijoles3, A. Sanchez-Valles1. 1) Morsani College of Medicine, University of South Florida, Tampa, FL USA; 2) Department of Pediatrics, University of South Florida Morsani College of Medicine, Tampa, FL USA; 3) Clinical Division, Greenwood Genetic Center, Greenwood, South Carolina USA.

We describe a 28-month girl with congenital hyperinsulinism/hyperammonemia (HH) syndrome due to a de novo mutation (His507Tyr) in the glutamate dehydrogenase gene (GLUD1) not previously reported in the literature. She initially presented with intracranial growth failure, foramen ovale, peripheral pulmonic stenosis, and dysmorphic features. Her initial exam revealed wide anterior fontanel, bilateral posterior hair whors, blue sclera, flat nasal bridge and posteriorly rotated low set ears. Chromosomal microarray analysis utilizing Affymetrix 6.0 platform revealed arr. (1-22, X) x2, with a high density of noncontiguous regions of homozygosity indicating possible consanguinity. Family history was unremarkable except that consanguinity was suspected as both parents were from a small village in Mexico. She had multiple short runs (1-10mb) of allele homozygosity throughout the genome, consistent with limited outbreeding. She was suspected to have a recessive syndrome that had not yet been identified. At 8 months old, she presented to the emergency room with new onset seizures and hypoglycemia (39 mg/dL). Seizures were described as episodes of unresponsiveness and staring, body stiffness occurred with some episodes and cyanosis with most. She had frequent seizures until 18 months of age, most of which were triggered by hypoglycemia. Electroencephalogram showed encephalopathy with epileptic activity in the left centro-parietal region. Her hypoammonemic episodes are triggered by dietary exigency. (6 µM) L at blood glucose of 44 mg/dL), negative urine ketones, low beta-hydroxybutyrate, and hyperammonemia (150-200 umol/L). DNA analysis revealed a novel mutation in GLUD1 that has not been previously described, predicted to be deleterious by SIFT and PolyPhen-2.0 programs. Parental DNA testing confirmed the mutation to be de novo, supporting the diagnosis of this autosomal dominant metabolic disorder. Her hypoglycemia initially responded well to 10 mg/kg/d diazoxide, however, she was subsequently admitted with recurrent seizures and gastroenteritis. Glycemic control was achieved with 15 mg/kg/d diazoxide and a leucine-restricted diet. At present, seizures are controlled with levetiracetam. She is thriving, but has global mild developmental delay that is attributed to her chronic hyperammonemia or an underlying unidentified genetic disorder whose locus is in one of the areas of homozygosity.

2395T
A novel mutation in GLUD1 causing Hyperinsulinism Hyperammonemia in a patient with high density of homozygosity on microarray. J.D. Odom1, M. Gieron-Korthals1, D. Shulman2, P. Newkirk1, E.J. Prijoles3, A. Sanchez-Valles1. 1) Morsani College of Medicine, University of South Florida, Tampa, FL USA; 2) Department of Pediatrics, University of South Florida Morsani College of Medicine, Tampa, FL USA; 3) Clinical Division, Greenwood Genetic Center, Greenwood, South Carolina USA.

We describe a 28-month girl with congenital hyperinsulinism/hyperammonemia (HH) syndrome due to a de novo mutation (His507Tyr) in the glutamate dehydrogenase gene (GLUD1) not previously reported in the literature. She initially presented with intracranial growth failure, foramen ovale, peripheral pulmonic stenosis, and dysmorphic features. Her initial exam revealed wide anterior fontanel, bilateral posterior hair whors, blue sclera, flat nasal bridge and posteriorly rotated low set ears. Chromosomal microarray analysis utilizing Affymetrix 6.0 platform revealed arr. (1-22, X) x2, with a high density of noncontiguous regions of homozygosity indicating possible consanguinity. Family history was unremarkable except that consanguinity was suspected as both parents were from a small village in Mexico. She had multiple short runs (1-10mb) of allele homozygosity throughout the genome, consistent with limited outbreeding. She was suspected to have a recessive syndrome that had not yet been identified. At 8 months old, she presented to the emergency room with new onset seizures and hypoglycemia (39 mg/dL). Seizures were described as episodes of unresponsiveness and staring, body stiffness occurred with some episodes and cyanosis with most. She had frequent seizures until 18 months of age, most of which were triggered by hypoglycemia. Electroencephalogram showed encephalopathy with epileptic activity in the left centro-parietal region. Her hypoammonemic episodes are triggered by dietary exigency. (6 µM) L at blood glucose of 44 mg/dL), negative urine ketones, low beta-hydroxybutyrate, and hyperammonemia (150-200 umol/L). DNA analysis revealed a novel mutation in GLUD1 that has not been previously described, predicted to be deleterious by SIFT and PolyPhen-2.0 programs. Parental DNA testing confirmed the mutation to be de novo, supporting the diagnosis of this autosomal dominant metabolic disorder. Her hypoglycemia initially responded well to 10 mg/kg/d diazoxide, however, she was subsequently admitted with recurrent seizures and gastroenteritis. Glycemic control was achieved with 15 mg/kg/d diazoxide and a leucine-restricted diet. At present, seizures are controlled with levetiracetam. She is thriving, but has global mild developmental delay that is attributed to her chronic hyperammonemia or an underlying unidentified genetic disorder whose locus is in one of the areas of homozygosity.

2396F
Pathogenic study of mitochondrial complex I deficiency and Leigh syndrome in Drosophila model. Z. Li1, K. Zhang2, J. Suan3, C. Haueter2, H. Sandovol1, H. Bellen1,2, B. Graham1. 1) Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX 77030; 2) Methodist Hospital, Houston, TX 77030; 3) Howard Hughes Medical Institute, Houston TX 77030; 4) Medical Student, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030.

Mitochondrial disorders, including oxidative phosphorylation (OXPHOS) disorders, affect at least 1 in 5,000 individuals. Isolated complex I deficiency is the most frequently observed OXPHOS defect (64%) and is the common cause of Leigh Syndrome with no effective treatments available. NDUFS3 is a Fe-S cluster-containing subunit of Complex I in electron transport chain. Oxidative stress is an important pathogenic factor in OXPHOS deficiencies as well as in diseases with secondary defects of mitochondrial metabolism. We have generated mutant NDUFS3 alleles in Drosophila through P-element excision of CG12079, the Drosophila ortholog of NDUFS3. Phenotypic characterizations include molecular, biochemical, immunohistochemical, electrophysiological and behavioral analyses as well as treatment with antioxidants. Reccessive mutations of NDUFS3 cause severe deficiency of Complex I activity with abnormal assembly. Most animals without NDUFS3 fail to pupate and die during larvalogenesis, however, rare adult escapers survive and experience a shortened lifespan. NDUFS3-deficient flies exhibit increased levels of oxidative stress demonstrated by fluorescence markers and decreased mitochondrial aconitase activity. Evidence for mitochondrial proliferation includes dramatically increased citrate synthase activity and increased Spargel gene expression (Drosophila ortholog of PGC-1 α). Mutant flies also exhibited neurological dysfunction manifested as increased sensitivity to mechanical stress, abnormal morphology of larval neuromuscular junction, defective climbing ability and perturbed retinal function. The adult mutant NDUFS3 phenotypes are fully rescued by ectopic genomic expression of NDUFS3 and partially rescued by vitamin E and coQ10 supplemented in food. These results demonstrate that the loss of NDUFS3 in Drosophila causes Complex I deficiency and increased oxidative stress. These studies also demonstrate that Drosophila is a useful model system for investigating the pathogenesis of mitochondrial disease as well as for exploring therapeutic strategies.
2397T
Long term follow up of 15 patients with methylmalonic acidemia following solid organ transplantation. J.L. Sloan, I. Manoli, C.P. Venditti. NIH/NHGRI, Bethesda, MD.

Methylmalonic acidemia (MMA) is a devastating disorder, recalcitrant to medical management in many cases. Solid organ transplantation, including liver (LT), combined liver-kidney (LKT) or kidney transplants (KT), has been reported to be associated with improved metabolic control. However, there is little systematic information on long-term outcomes and there are currently no guidelines regarding indications for transplant in MMA. We describe the natural history up to 15.4y post-transplant in 15 individuals with MMA (13 mut, 1 cblA, 1 cblB). There were 3 LT, 8 LKT and 4 KT recipients in our cohort. The mean age at transplant was 14.3 yrs (range 1.3-31.4 yrs) and mean follow up 5.1 yrs (range 0.8-15.4 yrs). All 11 patients status post LT or LKT remained metabolically stable (65.1 yrs combined follow-up), with the exception of one LKT patient who developed a severe gaseous distention and gastrointestinal dysmotility. A previously reported mut0 patient who developed renal failure requiring KT 1-year post-LT and a worsening movement disorder after liberalizing her diet. Another mut0 patient with KT suffered repeated hospitalizations for metabolic decompensations and developed optic nerve atrophy 3 yrs after transplant. Transplantation improved metabolic parameters (72-93% decrease), primarily due to correcting renal function in those who received KT, but all had persistently elevated methylmalonic acid in plasma (71-1383 µM, nl <0.4), urine (166-4446 mmol/mol cr, nl=3) and total whole body MMA output (23-382 µmol/kg/day). Despite the 8 MMA elevations, two LKT patients <12 yrs post transplant had eGFR >80 ml/min/1.73m2. Studies in this patient group, the largest assembled to date, afford the following conclusions: 1. LT and LKT, but not isolated KT, completely prevent MMA in all metabolic parameters. 2. After LT, patients can develop renal failure and therefore require careful monitoring and this is also a theoretical risk for LKT and KT patients. 3. Although solid organ transplantation improves biochemical parameters, adequate metabolic control is required following the procedure due to the long risk for late complications of the disease. Further studies are needed on the association of genotype and metabolic correlations pre- and post-procedure and the outcomes achieved, to develop guidelines about the optimal timing and procedure indicated for each patient.

2398F
Neuroimaging and neuropathology reveal dysmyelination in canine mucopolysaccharidosis I. P. Dickson1, J. Provenzale2-3, S. Chen4, I. Nestrasil1, S.-H. Kan5, S. Q. Le6, E. Lotshaw7, J. Jons8, J. Ye9, N.M. Ellinwood10, M.A. Guzman11, C. Vite4, V. Kovan12, E.G. Shapira13, 1 Dept Pediatr., LA Biomed Harbor-UCLA Med Ctr, Torrance, CA; 2 Duke University Department of Radiology, Durham, NC; 3 Emory University Department of Radiology, Oncology and Biomedical Engineering, Atlanta, GA; 4 University of Minnesota Department of Pediatrics, Minneapolis, MN; 5 Iowa State University Department of Animal Science, Ames, IA; 6 St. Louis University Department of Pathology, St. Louis, MO; 7 University of Pennsylvania, School of Veterinary Medicine, Philadelphia, PA.

The cause of neurological deterioration due to mucopolysaccharidosis (MPS) is poorly understood. While many investigations focus on neurodegeneration, evidence suggests that white matter involvement may also be important. Hyperintense white matter lesions on brain imaging studies correlate with cognitive impairment in patients (1), and MPS patients show reduced fractional anisotropy (FA, a measure of white matter integrity) on diffusion tensor imaging of the corpus callosum that correlates with specific measures of attention (2). We performed neuroimaging and neuropathological evaluation of the corpus callosum (a major white matter structure) in canine MPS I. We studied nine MPS I dogs, four unaffected carriers, and four MPS I dogs treated with intrathecal recombinant alpha-L-iduronidase (IT rhIDU) 0.05 mg/kg every 3m from 4m to 21m. Dogs were imaged at 4, 8, 12, 16 and 21 months and had IT rhIDU treated with Gb3 and/or DGJ.These experiments demonstrated that Gb3, possibly by interacting with TLR4, can induce a proinflammatory response (3). Moreover, addition of a TLR4-blocking antibody completely abolished the observed effect in M(+) DC and M(+) FDC and M(+) macrophages (M). The aim of this study is to investigate the effects of Gb3 on the production of proinflammatory cytokines in peripheral blood mononuclear cells (PBMC) from patients with EF, in particular dendritic cells (DC) and monocytes (M). The aim of this study is to investigate the effects of Gb3 on the production of cytokines in DC and macrophages (M) derived from normal PBMC monocytes treated with DGJ, an inhibitor of TLR4.

2399T
Neuropsychiatric Outcomes in PKU Patients Treated With Sapropterin: Results from the Randomized, Controlled PKU ASCEND (PKU 016) Trial. S. Prasad1, B. Burton2, A. Feigenbaum3, M. Grant4, R. Hendren5, R. Singh6, S. Stahl7, C. Zhang1, 1) BioMarin, Novato, CA; 2) Children’s Memorial Hospital of Chicago, Chicago, USA; 3) University California San Diego, San Diego, USA; 4) St. Christopher’s Hospital for Children, Philadelphia, USA; 5) University California San Francisco, San Francisco, USA; 6) Emory University, Decatur, USA.

Background: Phenylketonuria (PKU) patients often exhibit attention deficits similar to attention deficit hyperactivity disorder (ADHD). Methods: PKU 016 is the largest health outcomes study in PKU. This randomized, placebo-controlled trial evaluated baseline neuropsychiatric impairment, specifically ADHD behaviors, symptoms of anxiety (HAM-A rating scale), depression (HAM-D rating scale) and executive dysfunction (BRIEF rating scale), and the effects of sapropterin dihydrochloride (Kuvan) on these impairments after 13 weeks of treatment in 206 children and adults with PKU. The primary endpoint was the total score on the ADHD Rating Scale (ADHD-RS) commonly used to evaluate symptoms of inattentiveness and hyperactivity. Results: Mean patient age was 22.5±1.6 years, with 42% aged <18 years. Baseline blood Phe was 841.4±473.1 µmol/L. 35% of the sample had symptoms of ADHD at baseline. 118/206 were sapropterin responders (defined as blood Phe level reduction ≥20%). Among sapropterin responders, ADHD-RS total score (mean change from baseline -9.1±2.2 vs. -4.9±2.0, P=0.085), driven by a statistically significant and clinically relevant change in the inattention subscale (-5.9±1.4 vs. -2.5±1.3, P=0.036). Conclusions: Sapropterin was associated with improvement in inattention among children and adults with PKU who had a Phe response to sapropterin therapy.

2400F

Fabry disease (FD) is an X-linked genetic disorder characterized by the deficiency in the activity of the lysosomal enzyme a-galactosidase A (a-GalA). This defect leads to the accumulation of neutral glycolipids, mainly globotriaosylceramide (Gb3). Clinical manifestations in males include angioectasia, hypohidrosis, cornea verticillata, proteinuria, acroparesthesia. Death usually occurs at 40 decade of life because of kidney, cardiac or cerebrovascular problems. Previous results of our group showed increased expression and production of proinflammatory cytokines in peripheral blood mononuclear cells (PBMC) from patients with EF, in particular dendritic cells (DC) and monocytes (M). The aim of this study is to investigate the effects of Gb3 on the production of cytokines in DC and macrophages (M) derived from normal PBMC monocytes treated with DGJ, an inhibitor of a-GalA. PBMC were isolated from normal buffy coats, and M were purified and cultured in the presence of GM-CSF and IL-4, or M-CSF, to induce their differentiation into DC and MΦ, respectively. Obtained cells were then cultured in the presence or absence of 20µM Gb3 and/or 200µM DGJ and the levels of IL-6 and TNFα were analyzed in the supernatant. Both cultures showed a significant increase in production of IL-6 and TNFα, relative to the control, only in the case of combined Gb3 and DGJ treatment (DC p=0.0018 and p=0.0041, MΦ p=0.0002 and p=0.0054, respectively). A similar trend was observed for IL-6. Moreover, addition of a TLR4-blocking antibody completely abolished the observed effect in MΦ treated with Gb3 and DGJ. These results show that Gb3, possibly by interacting with TLR4, can induce a proinflammatory state, similar to that observed in PBMC from patients with FD, and could be directly involved in the pathogenesis of this disease.
Is common mutation in SC02 associated with reductive failure in the Polish population? Preliminary study. S. Luczak1, D. Piekutowska-Abramczuk1, A. Jezele-Stanek1, K. Chrzanowski1, M. Kugaudo1, A. Cieslikowska1, A. Kochanski2, P. Kowalski2, E. Ciara1, J. Trubicka1, D. Jurkiewicz1, M. Pelc1, M. Borucka-Mankiewicz1, M. Krajewska-Walasek1, E. Pronicka1. 1Medical Genetics, The Children’s Memorial Health Institute, Warsaw, Poland; 2Department of Child and Adolescent Psychiatry, Medical University of Warsaw, Warsaw, Poland.

Mutations in the SC02 gene (22q13) lead to severe COX deficiency observed mainly in muscles, heart, and brain. A common, g.1541G>A (p.E154K), substitution is identified at least on one allele (in heterozygous or homozygous form) in all reported patients. Analysis of genotype-phenotype correlation revealed the presence of two distinct clinical patterns of disease: phenotype of early beginning (at birth or in fetus) with hypertrophic cardiomyopathy as a leading symptom, and neurological phenotype of delayed beginning with encephalopathy. There are a few data of the possible association between mutation in SC02 and fetal wastage, especially in fetuses with a heterozygous form of common mutation. The aim of this preliminary study was to attempt to verify the existence of probable association SC02 mutations with early spontaneous abortions. We examined 64 DNA samples from couples with at least 3 miscarriages in family history (44 samples of whole blood on EDTA and 20 dry blood spots on Guthrie cards). Genomic DNA was eluted from dried blood spots using 5% Chelex-100 resin and extracted from peripheral blood leukocytes by phenol/chloroform procedure or by automatic isolation (MagNa Pure equipment). Genotyping for the presence of the specific mutation was performed by Real Time PCR analysis with Taqman probes (DNA concentration about 30 ng/μl). The probes were designed to 150 nt fragment of the SC02 gene (position in the reference sequence in the NCBI database: g:281182726:495-644). The study did not reveal the presence of common mutation in the examined group. Every patient was homozygous for the wild form. Since this is a preliminary study and the study group has less than 100 persons, the existence of correlation between reductive failure and SC02 mutations cannot be excluded. Carrierism of common mutations in the Polish population is relatively high (1:147), so the planned extension of the study group with early abortions seems feasible and justified. Implications: Weinstein M, McCullough KA, 11th Annual Meeting of the National Centre, EU Structural Funds, project POIG.02.01.00-14-059/09 and CMHI project no. S126/12.

Frequency of inherited metabolic diseases (IMDs) in high-risk children in a North Indian tertiary care hospital. S. Attri1, A. Patial1, P. Kumar1, P. Singh1, S. Singh1, S. Sharda1, C. Kumar2, I. Dwivedi1, S. Kapoor1, L. Kratzz1. 1Pediatrics, Postgraduate Inst Med Education & Research, Chandigarh, India; 2Division of Genetics, Lok Nayak Hospital & Maulana Azad Medical College, New Delhi, India; 3Biochemical Genetics Laboratory, KK, Baltimore, USA.

Introduction: Newborn screening, not commonplace in our country necessitates the shift to high risk screening. We present our data from a cohort in North India. Aims and objectives: To study the frequency of various IMDs in high-risk children. Methods: A total of 220 suspected cases with red flag signs suggestive of a metabolic disorder were enrolled from March 2012 till January 2013 from various wards of the Advanced Pediatric Centre like neonatal ICU, NICU, and the suspected diagnosis of SCAD was confirmed by finding increased plasma butyrylcarnitine level (C4: 1.92 nmol/ml; normal <0.62) and increased urinary excretion of ethylmalonate (EMA: 122 mg/g creatinine; normal <1). Sequence analysis of SCAD in the patients showed that the mother carried two apparent deleterious mutations (R107C and c.660-699del40) in cis and a presumed functional variant, g.281T, in trans; while the father was a carrier for a common variant, g.2029S. This case exemplifies the need for careful evaluation of parents of confirmed SCAD infants to identify subtle clinical symptoms suggestive of this recessive metabolic disorder.
**2405T**

The role of innate immune system activation and signaling in vascular disease in the MPS I canine model. M. Vera¹, S.Q. Le¹, S. Kan¹, P. Dickson¹, R. Wang². ¹Dept of Pediatrics, Division of Medical Genetics, Harbor-UCLA Medical Center, Torrance, CA; ²CHOC Children’s, Orange, CA.

Vascular disease in mucopolysaccharidosis I consists of regions of arterial intimal-medial thickening characterized by glycosaminoglycan storage, vascular smooth muscle proliferation, and elastin fiber disruption. The latter features are suggestive of innate immune system signaling pathways including those mediated by the Toll-like receptor family. These intraluminal lesions do not completely respond to enzyme replacement therapy or bone marrow transplantation, therefore understanding the initiating events leading to this phenotype is likely to be important in designing effective treatment. Our goal in this project is to identify and quantify the relationship between glycosaminoglycan storage and innate immune system mediated inflammatory signaling in MPS I vascular disease. Our hypothesis is that innate immune system stimulation leads to macrophage activation and production of cytokines, including TGFβ, that stimulate extracellular matrix remodeling and vascular smooth muscle cell proliferation. Our preliminary data indicate an increased expression of TGFβ and increased activation of downstream signaling mediators within these arterial lesions in MPS I canine arterial disease. We are continuing to study the causal relation between glycosaminoglycan storage and Toll-like receptor signaling in the development of this vascular phenotype.

**2406F**

Defective T-cell function in a mouse model of Citrullinemia Type I. T.N. Tarasenko, J. Gomez-Rodriguez, P.J. McGuire. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Although inborn errors of metabolism (IEM) are often exacerbated by infection, whether having an IEM impacts immune function is less well characterized. Lymphocyte activation is accompanied by dramatic changes in metabolism suggesting that certain IEM may also affect immune cell function. Argininosuccinate synthetase (ASS) is present in multiple tissues and is known to play a role in the conversion of citrulline to arginine, and nitric oxide generation. ASS deficiency in humans results in the urea cycle disorder known as known as citrullinemia type I, and is characterized by grossly elevated plasma citrulline levels and hyperammonemia. In human tissues, we demonstrated ASS protein in components of the immune system including naïve T-cells, thymus and spleen, however, the role of ASS in these normal tissues remains to be defined. Based on the expression of ASS, we hypothesized that ASS deficiency will result in T-cell defects. To evaluate this question, we are characterizing immune function in hypomorphic ASS¹/² mice. As previously published, the lifespan of fold/fold mice is reduced at 3 weeks. Interestingly, studies of immune organs revealed significantly smaller thymi and spleens, absent mesenteric lymph nodes, and a marked reduction in T-cell numbers despite normal expression of lineage and activation surface markers. Adeno-associated virus 8-mediated correction of liver ASS to enhance fold/fold survival resulted in a greater recovery of splenic T-cell numbers despite the persistence of the ASS enzyme defect in splenic T-cells. However, in vitro studies in fold/fold T-cells demonstrated abnormal differentiation into various T-cell subsets critical for immune responses including Th1, Th2, Th17 and Treg cells. In vivo characterization of fold/fold immune cells is currently underway using bone marrow chimeras with immunization and influenza infection protocols. In addition, a trans vivo approach using humanized mouse models is being utilized to explore the role of human ASS in immune organs and other immunologic niches. Finally, to further understand potential immune consequences of this IEM, clinical characterization of immune status in patients with ASS deficiency and related disorders is currently underway in the NIH Clinical Center. Overall, our work suggests that ASS deficiency may translate into specific immune defects and may have implications for patients with citrullinemia Type I.
2407W
Local and International Medical Geneticists and Genetic Counselors Collaboration: The First Vietnamese-North American Genetics Conference in Hanoi, Viet Nam. K. Leppard1, M. Laurino2, D. Sterren2, J. Thompson2. 1) Genetic Services, Group Health Cooperative, Seattle, WA; 2) Institute for Public Health Genetics, University of Washington, Seattle, WA; 3) Laboratory Medicine, Seattle Children's Hospital, Seattle, WA; 4) Molecular Genetics Laboratory, BC Children's and Women's Hospital, Vancouver, BC.

The commitment of local Vietnamese medical geneticists Dr. Dung Chi Vu at the National Hospital of Pediatrics and Dr. Nguyen Viet Nhan at Hanoi College of Medicine and Pharmacy served as the strong foundation for collaboration with international genetic professionals. The shared goals of the local and international partnership were to bring together genetic providers from across Viet Nam to discuss the scope of current genetics care provided in the country; facilitate collaboration and mentorship with international colleagues; discuss examples of new technologies that may allow more cost-effective genetic testing; and highlight the field of genetic counseling as part of health care delivery. A longer term goal was to develop a sustainable genetic counseling training program within Viet Nam to anticipate the need for specialized professionals who can work with the increasing number of geneticists as newborn screening and medical genetics services are rapidly growing. To achieve these goals, the first Vietnamese-North American Genetics Conference was held in Hanoi, Viet Nam on March 25-26, 2013. Targeted attendees of the meeting were health care providers in Viet Nam who are involved in the care of children with genetic conditions and birth defects. Over 300 physicians, laboratory specialists, nurses, researchers, medical students, international non-governmental organization workers, and government officials attended. There were 20 presentations by Vietnamese and international recognized North American geneticists. A genetic counseling workshop discussed the history of genetic counseling, roles of genetic counselors, establishment of the Philippines genetic counseling program as a model for strategic curriculum development, and the ethical, legal and sociocultural issues of genomic medicine. Conference attendees confirmed joint interest in the creation of a genetic counseling program within Viet Nam. As such, the 2nd Vietnamese-North American Genetics Conference is being planned in conjunction with the 2013 American College of Medical Genetics and Genomics conference in San Francisco, CA. The ongoing local and international partnership of genetic professionals. Certainly, this year's conference offered transparency and recognition of the current needs and invaluable opportunities on how to best integrate the field of medical genetics and genetic counseling in Viet Nam's existing health care and educational system.

2408W
Making sense of a primary care role in genetics: Views of genetics professionals. J.C. Carroll1, S. Morrison2, F.A. Miller2, B.J. Wilson2, J.A. Permaul2. 1) Dept of Family & Community Medicine, Mount Sinai Hospital, University of Toronto, Toronto, Canada; 2) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 3) Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; 4) Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada; 5) Department of Family Medicine, Mount Sinai Hospital, Toronto, ON, Canada; 6) Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada.

Background: Developments in genetics are widely expected to implicate primary care providers (PCPs), who, as first contact professionals with a foundational role in most health care systems, may need to be better equipped. Yet public policy efforts to enhance the relationship between primary care (PC) and genetics vary widely. Initiatives in some jurisdictions such as the UK seek to `mainstream' genetics knowledge and services, while in others such as Canada, top-down initiatives have not been pursued. Objectives: To explore the role of genetics health professionals (GHPs) in building relationships with the PC community. Method: Qualitative methodology. GHPs working at genetics centres in Ontario, Canada were invited to participate in semi-structured focus groups or telephone interviews. Purposive sampling was used to ensure diversity by type of GHP, and type of centre (academic, community, northern). Transcripts were analyzed using a qualitative interpretive approach. Findings: 5 focus groups and 2 interviews were conducted with GHPs (6 clinical geneticists, 24 genetic counselors, 4 nurses, 4 lab staff, 3 administrators) from 5 of 10 regional genetics centres and 2 of 3 northern outreach clinics across Ontario. Through interactions with PCPs, GHPs reported: very limited direct interactions with PCPs; and lack of education on how genetics might fit into PCPs' practice. GHPs saw a role for PCPs in genetics, but expressed concern at a perceived lack of knowledge of genetics, as reflected in both under and over referring, and PCPs' ability to incorporate new genetic discoveries into practice. GHPs also reflected on whether existing communication methods with PCPs were effective, and recognized a need to understand what PCPs might want from the genetics community. Conclusions: Widespread expectations of a PCP role in genetic medicine are shared by GHPs in Canada. But in the absence of policy direction or organizational support, GHPs struggle to make sense of how this might work. In this context, in addition to identifying knowledge and capacity deficits, parameters of a productive collaboration between PCPs and GHPs need to be developed.

2409W
A Brief Curriculum for Physician Orientation to Clinical Whole Genome Sequencing, M.A. Giovanni1,2, J. Krier1, J.L. Vassy1,2, D. Lautenbach1, M.F. Murray1,2. 1) Geisinger Health System, Danville, PA; 2) Harvard Medical School, Boston, MA; 3) Division of Genetics, Brigham and Women's Hospital, Boston, MA.

The MedSeq Project is a randomized clinical trial examining the impact of whole genome sequencing (WGS) in clinical care. In this study, 200 patients will receive WGS or usual care in the context of either primary or specialty care. Physicians are engaged by the study as research subjects and asked to recruit patients from their practice. This report describes a brief educational curriculum developed and administered to physicians participating in the research in order to prepare them to receive clinical reports from WGS. All of the physicians engaged in the educational program have appointments at Brigham and Women's Hospital in Boston. So far, we have enrolled 10 primary care providers trained in internal medicine (4 men and 6 women) and 7 specialists trained in both internal medicine and cardiology (5 men and 2 women). A 6.5 hour continuing medical education (CME) course was designed and offered to participating physicians prior to patient engagement. The curriculum included two 75-minute in-person didactic sessions, and 12 self-study cases, allowing multiple modalities to facilitate engagement of busy physicians. The case studies were designed to highlight key concepts in genomic medicine. In addition, a library of resources was created and made available online to supplement the course lectures and self study cases. Pre- and post-education surveys were conducted to measure the impact of the educational program. Additional data on the confidence and performance of physicians following the patient disclosures are being tracked. As genomic sequencing is integrated into the practice of medicine, motivated clinicians will seek opportunities to improve their knowledge and understanding of applied genomics prior to delivering results to their patients. A short curriculum of didactic sessions and self-study, coupled with relevant learning resources, has been created to enhance physician preparedness for delivering genomic results to their patients and data will be presented evaluating its effectiveness.

† Trainee Award Finalist
HelloGenetics (Alô Genética): A Brazilian strategy to provide guidance and education in genetics to primary health care providers.

Introduction: the integration of Medical Genetics (MG) into Primary Health Care (PHC) seems to be an alternative to develop actions of improved diagnosis, prevention and management of genetic conditions, and to facilitate access of patients to specialized health care. Therefore, it is desirable that PHC providers obtain basic knowledge about MG, including the immediate access of patients to specialized health care. Therefore, it is desirable that PHC providers obtain basic knowledge about MG, including the immediate access of patients to specialized health care.

Methods: a remote support facility was set up in 2013 at Medical Genetics Service, Hosp Clinicas Porto Alegre, Porto Alegre, RS, Brazil; 2) Bioethics Service, Research and Post-graduation Group, Hosp Clinicas Porto Alegre, Porto Alegre, RS, Brazil; 3) Department of Social Medicine, Universidade Federal do Rio Grande do Sul School of Medicine, Porto Alegre, Brazil; 4) InaGMP - Instituto Nacional de Genética Médica Populacional, Porto Alegre, Brazil; 5) Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Background: Due to new discoveries in genomic medicine, clinical guidelines are increasingly recommending the incorporation of genomic tests or therapeutics into routine care, especially as the transition of genetic knowledge from research laboratories into clinical practice is becoming more and more a part of health care systems. Primary care practitioners suffer from inadequate knowledge and skills in medical genetics and many are unaware of the technical, ethical, legal and psychosocial implications of genetic testing.

Methods: We initiated a "genomic education" program for the purpose of teaching primary care practitioners new advanced knowledge on genomic medicine. We emphasized the main take-home messages for physicians, which were: risk calculation for various genetic diseases, recognition of the mode of inheritance from the pedigree, guidelines for decision-making on which molecular tests to use, and the interpretation of test results on which a patient/family with a genetic condition, and to provide a long-term follow-up strategy to support their actions in the field.

Aims: to provide remote support to PHC providers to support their actions when facing a patient/family with a genetic condition, and to provide a long-range continuing education program in medical genetics to PHC providers.

Methods: a remote support facility was set up in 2013 at Medical Genetics Service, Hosp Clinicas Porto Alegre, in the Southern part of Brazil accessible by the toll-free telephone number 0800 642 6781 (in addition to email and internet). This new service was advertised at public health services of the Porto Alegre area by displays, and to a broader area also by emails. In addition, a continuing free education program in Portuguese was built, which could be accessed over the internet by PHC providers. Results: in the first months of activity, "Hello Genetics" was accessed by phone, email and internet by a growing number of PHC providers, mainly from the geographic area where advertisement was performed with displays; registration for the internet course was also being increasingly requested (courses to be started in Q3 2013), mostly from PHC providers from the same region.

Conclusions: the proposed strategy seems to fill a gap, helping PHC providers to access specialized medical genetics centers, obtaining the information needed and referring their patients, and improving the knowledge in the subject. However, advertisement should be increased to enable the program to expand nationally.

**2412W**


**2413W**

Human Genetic Variation: A Flipped Classroom Exercise in Cultural Competency. K. Tuttle, S. Dasgupta.

**2411W**

Hello Genetics (Alô Genética): A Brazilian strategy to provide guidance and education in genetics to primary health care providers.
2414W
How is high school students' genomic literacy correlated with their attitudes toward promotion of genomic studies? A comparison with a result of general public survey in Japan. 1. I. Ishiyama, 1. T. Tanzawa 1. 1. Department of Child Care and Education, Teikyo-Gakuen Junior College, Yamanashi, Japan; 2. Faculty of Education, Shizuoka University, Shizuoka, Japan.
This study aimed to 1) assess high school students' genomic literacy level, 2) reveal their attitudes toward genomic studies promotion related to medicine, and 3) identify associated factors to the approval of promotion. High school students were sampled from 2 high schools to be sure to contain a variety of academic students. The new self-reporting questionnaire was designed. An attitude survey was conducted in the 'Basic Biology' class at each school on January 2013. Students have already studied the unit of genetics. The date was analyzed using logistic regression models and compared with the results of nationwide opinion survey which was targeted general public and conducted in the same period. These surveys included the following same queries; 1) subjective and objective understanding of genome science, 2) pros and cons of medicine-related genetics studies promotion, and 3) attitudes toward science in general. Responses of subjective and objective understanding were calculated into scores of genomic literacy. Concerning high school students, the average score of subjective understanding was 8.6 (range 0-10), objective understanding was 5.3 (0-10), genomic literacy was total 13.5 (0-20). The average score of genomic literacy in general public was 10.5 (0-20). The term of 'Genome' was recognized by 99.7% students, while 31.9% of general public responded that they never heard of the term. Result also showed that approximately 3/4 students favored genomic studies related to medicine. Pros and cons were divided into agree (74.8%) and disagree (25.2%) categories; and cannot agree (24.4%), while only 0.8% expressed disagree. The similar result was shown in general public survey (agree: 71.6%, cannot agree: 27.6%, disagree: 0.9%). No correlation was found between literacy level and approval of promotion in students, whereas strong correlation was found in general public. Interest in science and technology in general, and willingness to use new products developed by technological innovation were correlated with the approval. In conclusion, even if high school students are knowledgeable about genomics, they don’t necessarily support the genomic studies in medicine. If anything, it’s important for them to develop their interest and trust in science through biology class.

2416W
Accelerating public awareness of personal genetics. M.E. Gelbart, L. Tomasselli, D. Waring, T. Wu. Personal Genetics Education Project, Department of Genetics, Harvard Medical School, Boston, MA.
The Personal Genetics Education Project (pgEd; www.pgged.org) is using a unique platform of strategies to address the widening gap between what researchers are learning at the frontiers of genetics and what the public understands. First, pgEd works to advance awareness through television programming with Hollywood, Health & Society at the USC Annenberg Norman Lear Center. Second, pgEd is continuing to develop the on-line game, Map-Ed (www.map-ed.org), which we launched in March to spread awareness of core concepts in genetics. Third, pgEd organizes the annual GE TEd conference, which brings together experts in education, research, health, entertainment, and policy to develop strategies for accelerating public awareness. Fourth, pgEd is building a freely available on-line curriculum that addresses the benefits as well as the ethical, legal, and social implications of personal genetics. Finally, pgEd is training high school teachers in biology, health, social studies, and beyond to engage an entire generation of young people on the topic of personal genetics.

2417W
Parents and newborn screening decisions: empirical assessment of specific educational messages. B.J. Wilson1, B.K. Potter2, J.C. Carroll2, J. Little3, D. Castle4, A. Avard1, P. Chakraborty2, S. Craigie1, H. Etchegary1, L. Lemery5, F.A. Miller1, G.A. Wells1, J. Millburn1, R. Rennicks White1, J. O’Sullivan6, G. Tawagi1, H. O’Sullivan1, G. J. L. Tawagi1.
1. Department of Epidemiology & Community Medicine, University of Ottawa, ON, Canada; 2. Family Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 3. Canada Research Chair in Human Genetics, Laboratory of Human Genetics, McGill University; 4. Centre for Human Genetic Resources, Irish Genome Centre, Dublin, Dublin, Ireland; 5. Centre of Genomics and Policy, McGill University, Montreal, QC, Canada; 6. Newborn Screening Ontario, Children’s Hospital of Eastern Ontario, Ottawa, ON, Canada; 7. Department of Pediatrics, University of Ottawa, ON, Canada; 8. Clinical Epidemiology, Memorial University of Newfoundland, St John’s, NL, Canada; 9. School of Psychology & Institute of Population Health, University of Ottawa, ON, Canada; 10. Institute of Health Policy, Management and Evaluation, University of Toronto, ON, Canada; 11. Department of Obstetrics, Gynecology and Neonatology, The Ottawa Hospital, ON, Canada; 12. Clinical Epidemiology Program, Ottawa Hospital Research Institute, ON, Canada; 13. Department of Obstetrics and Gynecology, University of Ottawa, ON, Canada.
Background: In many jurisdictions, systematic parental education about newborn screening (NBS) seems to be limited, perhaps because screening may be seen as routine care and not a separate intervention requiring conscious decision-making by a parent. However, even where screening is mandated, parental education provides benefits such as promoting trust in health services and preparing for the process and potential outcomes of NBS. However, NBS programs may be concerned that more comprehensive education may lead to higher parental anxiety and lower acceptance rates. Debates on secondary bloodspot use are also relevant to this issue. ‘Education’ is a broad concept, and there appears to be no empirical evidence on the most important content of education for parents. This study is designed to formally test the impact of specific components of potential NBS educational modules. The study aims to 1) measure knowledge and confidence among parents’ decision-making; 2) measure and compare parents’ acceptability of NBS and other reasons; and 3) identify associated factors to the approval of promotion.
Methods: The study sample is women with low risk pregnancies attending routine screening at three Ontario birthing hospitals. Using a factorial design (objectives: 2; knowledge: 2; confidence: 3), 36 combinations of messages (cognitive burden), and associations with other respondent characteristics were calculated. Participants were randomized to one of the 36 combinations of messages (factorial design). The primary outcome is the mean score (max 100) on the Decisional Conflict Scale (DCS). Results: From a target sample size of 500, we report data on the first 129 respondents. The mean DCS score is 27 (range 0-89). Over 90 percent intend to accept NBS; actual intentions are likely to be even higher due to prospect theory. In univariate analyses, no statistically significant differences are observed in DCS scores in those receiving/not receiving specific messages, with the exception of the storage/secondary use of bloodspots message (p=0.02), a lower mean DCS score was observed. First, pgEd is working to advance awareness through television programming with Hollywood, Health & Society at the USC Annenberg Norman Lear Center. Second, pgEd is continuing to develop the on-line game, Map-Ed (www.map-ed.org), which we launched in March to spread awareness of core concepts in genetics. Third, pgEd organizes the annual G E TEd conference, which brings together experts in education, research, health, entertainment, and policy to develop strategies for accelerating public awareness. Fourth, pgEd is building a freely available on-line curriculum that addresses the benefits as well as the ethical, legal, and social implications of personal genetics. Finally, pgEd is training high school teachers in biology, health, social studies, and beyond to engage an entire generation of young people on the topic of personal genetics.
2418W
On 30th anniversary of Orphan Drug Act: a review of resources to inform and connect rare disease patients. M. H. Dunkle1, J. O’Leary2, H. Hyatt-Kornell2, M. Snyder4, J. Lewis3. 1) Communications, National Organization for Rare Disorders (NORD), Danbury, CT; 2) Genetic Alliance, Bethesda, MD; 3) National Institutes of Health Genetic and Rare Diseases Information Center, Bethesda, MD.

The Orphan Drug Act (ODA) of 1983 brought major new incentives to encourage the development of therapies for patients with rare diseases. It also launched a coordinated effort, involving patient organizations, government agencies and others to address the unique and previously unaddressed needs of patients and families affected by rare diseases. During this 30th anniversary year, three of the primary sources of information and connection for rare disease patients provide a brief review of the history of rare disease information since 1983 and highlight a few of the major resources available to patients and families today. The three collaborating entities are the National Organization for Rare Disorders (NORD), which was established by patient organizations in 1983 after the ODA was enacted; Genetic Alliance, established in 1986 by NORD, the March of Dimes, and the Maternal and Child Health Bureau of the Health Resources and Services Administration; and the NIH Genetic and Rare Diseases Information Center (GARD), established in 2002 by the NIH National Center for Advancing Translational Sciences/Office of Rare Diseases Research and the NIH National Human Genome Research Institute. Current initiatives include NORD’s Rare Disease Database, regional patient organization meetings, and RareConnection, global online patient communities established by NORD and its European counterpart, EURORDIS; Disease InfoSearch, an online database of advocacy organizations and resources for genetic conditions and Genes in Life, to provide easy access to Genetic Alliance’s many resources and tools related to health and genetics information; and the GARD Information Center for patients and their families, as well as researchers and medical professionals, in which trained Information Specialists staff a toll-free phone line, provide customized responses in English and Spanish, and create web content based on questions received from the public so that it is responsive to public 'here and now' issues. Together, these three entities, NORD, Genetic Alliance, and GARD, are providing valuable and complementary information services to the rare disease community.

2419W

We stand today on the threshold of a new era. Within the foreseeable future analysis of complete genomes will play a prominent role in society in the broadest sense. It will have immediate implications for healthcare, agriculture, biodiversity research and many other aspects of the world around us. The genomic era is a Dutch initiative that aims to create public awareness of DNA-sequencing and its applications. The focus is on high school students, as they already possess some background knowledge and their generation is thought to be exposed to DNA sequencing most frequently in the near future. Our approach is to invite the public to interesting questions that can be addressed by DNA sequencing. LeveDNA! attempts to realize such ideas by finding support for sampling, sequencing and analysis. Secondly, we actively initiate projects that bring our target audience in contact with DNA technology in a hands-on way. At this point we report on projects that have been conducted so far include: (I) Sequencing the genome of a DJ from a popular Dutch radio station, followed by a dedicated evening of his daily radio show. During the three-hour broadcast, listeners were given the opportunity to ask questions about the DJ’s genome. Is he a carrier of a genetic disorder? Will he be bald? Is he a night owl? Are some of the many questions asked. (II) LeveDNA!, in collaboration with an initiative to teach the public about evolution (Evolution MegaLab), made an attempt to identify genes that contribute to the colour patterning of garden snail shells. (III) Organisation of a DNA-day, where participants are taught about DNA and where they contribute to the colour patterning of garden snail shells. (III) Organisation of a DNA-day, where participants are taught about DNA and where they contribute to the colour patterning of garden snail shells. (IV) Shaking Science, a biodiversity project where students sampled fresh water around the Natural History museum; the sequence determined can be analysed by everybody using a smartphone. Finally, our high school students are running organisms in fresh water of a major river in the Netherlands, the Rhine. LeveDNA! creates a platform where the general public and experts directly interact, exchange knowledge and discuss social consequence of the new technologies. LeveDNA!: ‘Leve’ (Dutch) stands for both ‘life’ and ‘hurray’. 2421W
Vignettes as an aid to deciding about genetic testing. D. Zallen. Dept Sci & Tech in Society, Virginia Tech, Blacksburg, VA.

Genetic tests are currently offered in doctors’ offices and by commercial laboratories widely advertised on the Internet. Though genetic information can be very useful for some people, for others it can produce severe personal and family problems without any significant corresponding medical benefit. Within the confines of a standard medical office visit, it is difficult to provide consumers with the information that will enable them to make informed decisions about whether or not to test. New educational approaches are needed. Consumers deciding whether or not to test need guidance not only in understanding the technical aspects of genetic testing but also in assessing its suitability with regard to their personal values, psychological makeup, and family dynamics.

As part of an investigation of new educational tools, we have developed a set of short dramatized vignettes based on more than 150 interviews (D.T. Zallen, “To Test or Not to Test”, Rutgers 2008) with consumers who were relying solely on written text and were an effective way to help them identify their own value preferences and reach appropriate decisions. The use of dramatized vignettes is a novel and effective approach to providing consumers with a means of clarifying the various value dimensions associated with genetic testing prior to deciding whether or not to undergo a genetic test.
**2422W**

Incorporating Bioinformatics into the Undergraduate Genetetics Curriculum through an Authentic Research Project. B.V. Bowling, E.D. Strome.

Genomics and bioinformatics are advancing fields that have become an integral part of a broad spectrum of biological research. As a result, it is increasingly important for undergraduate genetics courses to include basic bioinformatics tools and genomic concepts to prepare students for the modern biology workforce. At Northern Kentucky University, we’ve introduced these through an authentic research project in which students investigate the role of *Saccharomyces cerevisiae* open reading frames designated as coding for ‘putative’ elements within the Saccharomyces Genome Database. Students create knockout strains for their gene of interest and then analyze the phenotypes, alongside bioinformatics analyses conducted on individually assigned ORFs and then data synthesized in order to develop a hypothesis of the function of the gene product within the cell. The bioinformatics and comparative genomics analyses include multiple sequence alignment, conserved domain identification, cellular localization prediction, reviewing interaction and expression data, and gene deletion phenotypes. The *in silico* approaches allow students to apply genetic (and additional biological) concepts to an authentic research question, provide experience with bioinformatics tools, and challenge students to analyze and synthesize data from various sources. The project culminates in a brief presentation by each student in which they summarize the most important evidence collected for the role of their gene product in the cell. These activities have been shown to be effective in our sophomore-level, multi-section laboratory course involving several instructors. Instructors report increased engagement relative to previous labs, with a substantial portion of the students (74%) felt that they learned even more from this project than other lab projects. The vast majority of the students indicate the research was challenging (93%) and that they are better prepared to take on investigating a problem without a known solution (82%). The lack of established answers contributed to student anxiety about the assignment. Many students have the urge to check to see if they are ‘right,’ and instructors have to balance encouraging independence while also reassuring students of their analyses. Incorporating formal mechanisms for peer feedback throughout the process has helped with students confidence in their work.

**2423W**

Integrating authentic research in human genetics into an undergraduate liberal arts curriculum to enhance the pace of rare disease gene discovery for underserved populations. R. Jinks1,2 B. Davis3, D. Roberts1, E. Rice1, K. Brigatti1, E. Puffenberger4, K. Strauss3,4. 1) Biology, Franklin & Marshall College, Lancaster, PA; 2) Biological Foundations of Behavior Program, Franklin & Marshall College, Lancaster, PA; 3) Clinic for Special Children, Strasburg, PA; 4) Lancaster General Hospital, Lancaster, PA.

Next-generation sequencing accelerates disease gene discovery, especially for orphan diseases, though at present it outpaces functional studies needed to provide ‘proof of causation.’ This problem was underscored by a recent funding opportunity from the NHGRI requesting assistance with functional studies for 75 rare, potentially pathogenic alleles discovered through the NHGRI Undiagnosed Diseases Program. Working from a small, community-based, clinical laboratory, the Clinic for Special Children has identified more than 130 allelic variants associated with disability, disease, or untimely death among the Amish and Mennonite (Plain) populations of North America. In collaboration with the Clinic, we developed an HHMI-funded program that integrates rare disease gene discovery research into our undergraduate curriculum. Roughly 150 students per year in our introductory cell biology and neuroscience courses trace disease genes and study the functional impacts of gene variants through expression in mammalian cell culture. We use these authentic research experiences to teach key concepts in cell biology, genetics, and neuroscience. Students later build upon this experience in upper-level courses in neuroscience, genetics, cancer biology, and immunology in which they engage in semester-long research projects in small teams, each conducting functional studies of a single novel gene variant. These experiences engender talented undergraduates to assume greater research responsibility through independent study and summer research projects that transition them into the role of co-PI for the gene or disease they have chosen to study. Data published through this project (PLoS ONE 7:e28936) were recently used by another institution to diagnose a critically-ill, non-Plain newborn with lethal neonatal seizure-rigidity syndrome (Sci Transl Med 4:154ra135). This provides important proof-of-concept for integrating novel disease gene functional studies into a carefully structured undergraduate research curriculum. Students in our Public Health program are collaborating with the Clinic to develop handbooks to help parents care for children with special medical needs with the goal of producing one to two high-quality disease handbooks annually. Our program represents a model for engaging undergraduates in meaningful research at the front lines of biomedical science and public health in ways that directly impact the diagnosis and care of children with rare inherited disorders.
2425W
The Research Connection: Development of an integrated institutional pediatric research infrastructure as a model framework for multi-institution implementation and collaboration. S.K. Savage1, C.M. Clinton1, W.A. Wolf2, D.M. Margulies1,2,3,4. 1) Program in Genomics, Boston Children’s Hospital, Boston, MA; 2) Department of Pediatrics, Harvard Medical School, Boston, MA; 3) Division of Developmental Medicine, Boston Children’s Hospital, Boston, MA; 4) Center for Biomedical Informatics, Harvard Medical School, Boston, MA.

Our objective is the development and implementation of a fully-integrated research infrastructure to power and enhance discovery within Boston Children’s Hospital (BCH) and across collaborating institutions. With our ultimate goal, we reviewed and assessed existing research resources and identified gaps and areas of need. We then designed a set of core research and clinical services, called the Research Connection, to supplement existing operational structures to meet varied needs of BCH investigators, clinicians, and collaborators. The BCH Biobank is patient-centric repository of leftover clinical samples linked with electronic medical record data to facilitate biospecimen-based research. By offering Biobank enrollment across the entire institution, we attain broad reach over a diverse population of children and adults, creating a large and rich research cohort. Claritas Genomics, BCH’s new genomics and genomics diagnostic laboratory, provides cutting-edge measurements, analyses, and revisional reporting for both research and clinical samples. BCH’s Interpretive Genomic services allow for refined interpretation, clinical correlation, and specialty care management, drawing on clinical and scientific expertise within the institution. Finally, our novel approach to integrated clinical and research genetic result leverages existing and expanding BCH clinical services to provide appropriate result disclosure. Our Research Connection structure to be optimized to accommodate multi-institutional collaboration via a new collaborative agreement structure called the Pediatric Research Network. To support the concept of a Pediatric Research Network, we have initiated collaborative relationships with other pediatric institutions, where components or the entire Research Connection model may be adopted and tailored to meet their unique institutional needs. By implementing a common research infrastructure like the Research Connection across multiple sites, coupled with a collaborative agreement structure, we envision the opportunity for an extensive ‘virtual’ research repository to support large-scale research, foster collaborations, and enhance discovery that would otherwise not be possible. This structure may be the future of genomic research design for large and small institutions.

2427W
Participant views of re-consent and broad consent in cancer genetics research. K. Edwards1,2, D. Konigstiel3, L. Mepier2, J. Scott3, N. Shridhar4, D. Kaufman4, D. Bowen4, C. Condit1. 1) Institute for Public Health Genetics, School of Public Health, University of Washington, Seattle, WA; 2) Epidemiology, School of Public Health, University of Washington, Seattle, WA; 3) Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA; 4) National Coalition for Health Professional Education in Genetics, Lutherville, MD; 5) Genetics & Public Policy Center, Johns Hopkins University, Washington, DC; 6) Community Health Sciences, School of Public Health, Boston University, Boston, MA; 7) Communication Studies, University of Georgia, Athens, GA.

Background: Institutional Review Boards and researchers protect participants in genetic research, but the preferences of participants do not sufficiently inform policy currently. With the proliferation of biobanks intended to support future studies that go beyond the scope of their original consent, fully understanding participant preferences for re-consent is a pressing concern. The Participant Issues Project (PIP) addresses this gap. Methods: PIP study participants were drawn from the Northwest Cancer Genetics Registry and included cancer patients, controls, and relatives. Thirty telephone interviews were conducted and analyzed using content and thematic analysis. Questions asked participants to consider diverse scenarios: the research focus had changed from the original study; re-consent of a minor achieving majority; for-profit studies; and the use of broad consent forms. Results: The majority of participants agreed that re-consent was necessary when the study direction changed significantly (e.g., a different disease) or a child participant became an adult. Participants favored re-consent in order to provide study information and control of the use of their data. However, participants often considered their consent to be primarily for research and not for treatment purposes. Conclusions: There were specific scenarios where the majority of participants felt re-consent was acceptable. Future research should focus on developing best practices for future research and broader consent forms. The perceived value of re-consent appears to lie largely in keeping participants informed of future studies and updated on the research progress, while for others, it is to control the use of their data. It is not known whether satisfying the need for information will change participant perspectives on the desire for re-consent.

2426T
Revisiting the role of gamete and embryo donor registries on the transmission of genetic information between donation relatives, in light of the principle of beneficence. V. Couture1, M.-A. Dubois1,2, R. Drouin1, J.-M. Moutquin2, C. Bouffard1. 1) Division of Genetics, Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Quebec, Canada; 2) Me Hélène Guay Office, Montreal, Quebec, Canada.

Introduction: In the context of assisted reproductive technology, the transmission of genetic information between relatives may have implications for prevention, diagnosis, treatment, and reproductive decision-making. In the case of gamete and embryo donor-conceived people, where there is no previous contact between donors and their sibling, this issue acquires greater complexity. The creation of gamete and embryo donor registries is one of the strategies that has been developed to make information available between donation parties. However, the nature, flow, preservation, and accessibility of genetic and medical information depend on the type of registry. Whether national, institutional or independent, each type may have different repercussions on the health protection of donor-conceived people, donor, siblings, and offspring. Under these conditions, our objective was to identify the ethical implications of the transmission of genetic and medical information through donor registries, in light of the principle of beneficence. Methods: Bioethical analytical approach: (1) typology of donor registries (2), analysis of their strengths and weaknesses, and (3) introduction of the principle of beneficence in the analysis of their impact. Results: A donor registry has as its main benefits the opportunity to track hereditary disease and prevents future gamete donation by an affected donor. It can also help in the diagnosis of genetic disease as well as in the prevention of late-onset conditions. According to the principle of beneficence, the needs and values of all donation parties have to be considered. A registry must address the issues of genetic and medical information, as well as the protection of their anonymity. An effective counseling service has to be provided to avoid unnecessary distress for all the parties. To maximize the benefits of gamete and embryo donor registry, the national type with proper counseling and clear guidelines appears to be the most respectful solution to promote the health of all donation parties. Conclusion: Despite the complexity of genetic and genomic issues for the development of a national donor registry, the principle of beneficence offers us the opportunity to pursue the present and future needs of all donation parties.
2428T
Balancing Patient Privacy While Supporting Rich and Convenient Access to Clinical Genomic Data. E.R. Riggs1, J. Berg2, C. Bustamante3, D.M. Church4, W.A. Fauquet5, M. Feolo6, D.H. Ledbetter7, D. Magiorakos8, C.L. Martin9, D. Metterville6, J. Mitchell2, R. Nussbaum10, J. Osteli11, S. Plon5, H. Rehm5,6, L. Rodriguez10, W. Rubinstein10, M. Watson11, International Collaboration for Clinical Genomics. 1) Geisinger Health System, Danville, PA; 2) University of North Carolina, Chapel Hill, Chapel Hill, NC; 3) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 4) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD; 5) Laboratory for Molecular Medicine, Partners Healthcare Center for Personalized Genetic Medicine, Cambridge, MA; 6) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT; 7) Division of Genomic Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA; 8) Departments of Pediatrics and Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 9) Department of Pathology, Harvard Medical School, Boston, MA; 10) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 11) American College of Medical Genetics and Genomics, Bethesda, MD.

It is becoming clear that access to large datasets of genomic variants and observed phenotypic consequences is critical to our understanding of human genomic variation and its applications to basic science, public health, and improved patient care. However, these datasets must be maintained in ways that both respect patient privacy and support active use by the scientific community. The International Collaboration for Clinical Genomics (ICCG) is a group of laboratories, clinicians, and researchers working closely with the National Center for Biotechnology Information (NCBI) to develop ClinVar, a public database of clinically relevant genetic variants, associated phenotypic information, and clinical classifications with supporting evidence. Laboratories are beginning to submit their data, which include a mixture of both de-identified and potentially identifiable information regarding variants and individual cases observed during the course of clinical testing and research studies. Given that a large amount of ClinVar’s data will be derived from patient testing results and may warrant different levels of protection, it is critical that ICCG develops clear policies regarding patient privacy, and clinical data access. Current data access policies and requirements of similar databases were initially developed to accommodate traditional research applications of the data, and remain appropriate for these types of applications; however, many of the requirements, such as IRB-approved protocols, may not be appropriate for the inclusion of public data. How can we balance data access and confidentiality of the patients?

2430T
NIH’s approach and expectations for genomic data sharing: The draft NIH Genomic Data Sharing Policy. D. Pattlo1, A. Bailey1, C. Fomous2, K. Langlais2, E. Lueketmeier3, T. Payne1, R. Wise3, L. Lyman Rodriguez2, 1) Genetics, and Health, Society Program, Office of Science Policy, Office of the Director, NIH, Bethesda, MD; 2) Division of Policy, Communications and Education, National Human Genome Research Institute, NIH, Bethesda, MD.

Since 2007, the Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) has governed the sharing of GWAS data from human participants in research supported by NIH. The GWAS Policy established a presumption that all such data would be made available, with exceptions as warranted, for secondary research uses that are consistent with research participants’ informed consent. Rapid advances in DNA sequencing and other high-throughput technologies have increased the volume, complexity, and types of data generated in genomic studies. To ensure the full value of genomic data in light of this growth, NIH has drafted the Policy for Genomic Data Sharing in NIH-Supported or NIH-Conducted Studies (GDS Policy), which updates and expands on the 2007 NIH GWAS Policy. The draft GDS Policy encompasses a broader range of data types (e.g., GWAS data and genomic, transcriptomic, or epigenomic data produced by array-based technologies, or data obtained from high-throughput sequencing technologies) from research in both human and nonhuman or model organisms, and updates expectations for data submission, data release, and human participant protections. As with the NIH GWAS Policy, the draft NIH GDS Policy is founded on the principle of maximizing public benefit by facilitating broad data sharing to advance the understanding of public health needs, while ensuring the responsible oversight of data that may contain de-identified and potentially identifiable information. The NIH Guide as well as two public webinars in the summer of 2013, NIH obtained public perspectives on all aspects of the draft GDS Policy, as well as on related data management topics such as the potential benefits and harms of broadly sharing genomic research data; scientific, technological, policy, or legal strategies or factors, that would minimize risks and should be considered when genomic data is shared through open access repositories; and models that NIH should consider to facilitate access to large-scale complex datasets. NIH also considered data use policies, and expectations for data integrity. The comments were publicly posted on the NIH GDS website (http://gds.nih.gov) after the close of the public comment period. NIH expects this policy to be implemented in 2014. This presentation provides an overview of the draft NIH GDS Policy, the public consultation process and NIH's approach and expectations for NIH GDS Policy implementation.

2431W
The Potential Power of Personal Genomics in Reducing Social Stereotypes: Attitudinal Study and Computer Animation of Results for 4,000 Japanese Respondents. T. Kido1, M. Swan2. 1) Rikenkensai, Tokyo, Japan; 2) DIYgenomics, Palo Alto, U.S.

This paper provides a flagship study of ethics and discrimination as individuals learn information about their genome. The study reports on the results of the analysis of the social psychological statistics related to the impacts of the emergence of personal genome services. We were interested in determining how we might choose to respect others, not discriminating against them, while we believe that they are not discriminating against us. We would also like to understand the basic mechanisms underlying the reason for why discrimination emerges in our society. We conducted focused group discussions in a small lecture group in Tokyo University. Next, based on these discussions, we developed web-based survey systems with computer animations. We conducted 30 web-based interviews with 30 to more than 4000 Japanese individuals. A Focus Group Discussions Focus group discussions on personal genome services revealed that a variety of opinions exist simultaneously by both positive and negative sides. We classified these opinions, and asked more than 4000 Japanese people to vote on their favorite opinions. B. Web-based surveys and statistical investigations We created animation stories for symbolizing discrimination against people with a handicap and analyzed the responses. Our preliminary results that involved survey with more than 4000 volunteers revealed the following: 56.9% of people believed that society discriminates against people, with handicaps. 61.7% of people wished to have information about their genomes, even if it revealed a handicap. These preliminary results pose interesting questions about why a majority of the Japanese people might believe that discrimination exists against them, while they believe that they themselves are not discriminating. Social psychological research is expected to shed further light on these questions. This work demonstrates the importance of constructing frameworks for philosophical, psychological, and technological research on personal genome information technologies. One of the biggest things we have seen emerge as people have their genetic data is that everyone realizes they have one ‘handicap’ or another - everyone is at higher genetic risk than the average for one of the top 20 health conditions. The GWAS Project does this to the extreme. A: Me (non-handicapped) vs. them (handicapped): B. We are all ‘handicapped’ (genetically predisposed) for some condition This has been one of the most great destigmatizing influences of genomics.
2432T


PURPOSE: The aim of this study is to describe the hopes and expectations of individuals engaged in the process of utilizing next generation sequencing (NGS) as a clinical tool. Participants include patients presenting to the Mayo Clinic Individualized Medicine Clinic (IM Clinic) who have a cancer that has failed standard treatments or have a diagnostic odyssey, their family members, and the clinicians involved in the care of the patient. METHODS: We conducted semi-structured interviews with 16 patients and their family members who were referred to the IM clinic throughout multiple points of the process of NGS. Audio recordings of the interviews were transcribed and analyzed using inductive narrative analysis. Future research activities will include more interviews and surveys of patients and their families entering the IM Clinic and throughout their journey as well as clinicians. RESULTS: The patients verbalized their understanding that the probabilities of finding something from NGS that will benefit them directly are low. However, patients maintained optimism despite having been advised by clinicians and genetic counselors that NGS is a relatively new process in the clinical realm and as such the possibility of it resulting in a beneficial outcome cannot be guaranteed. A prominent theme that emerged from the interview data was that the patients’ hopes were nuanced by more ‘realistic’ expectations. To illustrate, one participant said that the hopes were to find out anything possible and her mother added: ‘Our hopes are that...next week you will call and say we found a gene and we can fix it’ to which the patient inserted a sarcastic aside that demonstrated her awareness of the likelihood of their hopes coming to fruition: ‘And we’re buying you Cinderella’s castle and you get to live in it! But despite their tempered understanding, many patients upheld what they knew to be unrealistic hopes and chose to move forward with NGS anyway. DISCUSSION: Patients’ hopes and expectations can influence their decision-making in deciding whether to proceed with novel clinical testing. Understanding the interplay of patients’ hopes and expectations may prove valuable to health care providers and genetic counselors as they guide individuals through the process of NGS. As NGS becomes increasingly integrated into clinical care, clinicians and researchers must continue to investigate the implications of using these technologies from a patient perspective.

2433W

Genetic privacy in the European Union - exploring the impact of the proposed Data Protection Regulations. A. de Paor. Centre Disability Law & Policy, Natl Univ Ireland, Galway, Galway, Galway, Ireland.

With advances in genetic science and technology, genetic information is becoming increasingly available and accessible. However, the deluge of genetic information raises many ethical and legal issues that may threaten advancing science, in the absence of appropriate regulation. One such issue is the violation of genetic privacy. Genetic information is a sensitive and inherently personal type of information that can reveal intimate details about an individual and an individual’s family. In the nature of genetic information, the potential abuse is clear, as is the desire to protect such information from access and disclosure. Therefore, there are compelling reasons to maintain the privacy and confidentiality of genetic information. In the European Union (EU), the Data Protection Directive (introduced in 1995) provides a strong data protection framework that protects the privacy of personal information. However, there is currently no substantive provision for genetic data and no specific reference to this category of information, intangible medical/medical genetic data. The Directive provides no guidance for genetic data in the EU. Recent developments indicate an intention to include genetic data within the scope of EU data protection laws, with the European Commission proposing a new Regulation. With these new Regulations, the Commission aims to develop an updated data protection framework. It is committed to reform and modernize data protection legislation, in line with the realities of today’s society, and changing norms. The draft Regulations identify ‘genetic data’ as a category of personal data designated for special protection. ‘Genetic data’ is defined broadly to include all data, of whatever type, concerning the characteristics of an individual’s or an individual’s family’s genetic constitution or their genetic state, thus presumably incorporating all genetic data as well as family medical history and other genetic related health information. As regards protection of genetic privacy in third party contexts, these draft Regulations are welcomed for their intention to recognize the inherent nature of genetic data that deserves protection. This paper will explore the issue of genetic privacy in the EU. It will examine the current data protection framework, as well as the proposed reform of the EU data protection framework and the impact that these reforms will have on the protection of genetic data.

2434T

An evidence-based framework for incidental findings from exome sequencing in the pediatric setting. E.T. DeChene1, 2, S. Mulchandani1, M.C. Oulkil1, 2, 3, 4, L.K. Conlin1, J.L. Abruend1, 2, B.A. Bernhardt1, K. Izumi1, S.E. Noon1, R.E. Pyeritz2, A. Santarini1, J.F. Slack1, C.A. Stolze1, A.B. Wilkens2, I.D. Krantz1, N.B. Spinner2, 1) Dept of Pediatrics, Children’s Hospital of Philadelphia, Philadelphia, PA; 2) Dept of Pathology & Laboratory Medicine, Children’s Hospital of Philadelphia, Philadelphia, PA; 3) Dept of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Clinicians and laboratories are considering how much secondary information to return to patients undergoing genome-wide tests. The use of exome sequencing, in particular, has increased in recent years secondary to incidental findings (IFs), which are clinically relevant genetic variants unrelated to the reason the test was ordered. The issues surrounding genetic IFs in pediatrics are particularly complex, because most children and adolescents cannot fully comprehend implications of potential IFs or independently consent to testing. However, ACMG recommends that laboratories search out and report certain medically actionable IFs, due to the potential impact on the child’s or family’s health. Funded by the NHGRI’s Clinical Sequencing Exploratory Research Program, the Pediatric Genetic Sequencing (PediSeq) Project at The Children’s Hospital of Philadelphia (CHOP) and University of Pennsylvania is exploring best methods for performing exome and genome sequencing and returning IFs in the pediatric setting. PediSeq’s Oversight Committee (OC) carefully considered all result options centered to families, creating a balance in the principles of beneficence, non-maleficence and patient autonomy. Our current framework allows families to opt to receive results related to medically actionable (MA) childhood-onset disease, MA adult-onset disease, and carrier status for recessive disorders. The OC expected little change in treatment or care based purely on genetic diagnosis (in the absence of symptoms). We believe that we are obligated to release MA results regardless of a family’s preferences, due to the likely impact on a child’s long-term health. Initially categorization of IFs was established by OC consensus. Due to the subjectivity of this approach, we developed an evidence-based categorization method centered on gene/disease evidence; affected gender and age of onset; availability of treatment or management; and other considerations. Here, we present our current IF framework and verify its utility by categorizing over 40 IFs identified in the CHOP Cyogenetics Laboratory. For example, Brugada and several childhood cancer syndromes were categorized as MA. We propose that carefully categorizing IFs and offering families options for medically actionable IFs when possible, is the best way to balance patient autonomy with our ethical and legal responsibilities to patient care.
2436T Research Policy of the Genome Science Project in Japan. J. Minari1, K. Kato2, 3. 1) Biomedical Ethics and Public Policy, Osaka University, Osaka, Japan; 2) Institute for Integrated Cell-Material Science (iCeMS), Kyoto University, Kyoto, Japan.

A new large-scale project, the Genome Science Project, started in Japan in 2010. It is a five-year project funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The project aims to support various genome researchers in Japan who are funded by MEXT and selected through an annual open call; it does so by providing services of high-throughput DNA sequencing and high-grade information technology. A program of the Genome Science Project is the Genome Medical Group (GMG), which collects human specimens and sequences the whole genome/exome to support the selected researchers, and thus naturally has to take into account various ethical, legal, and social implications. For this, a special group, Research Unit for Genome Ethics (GenEth), has been established within the project. In starting the MGSP, we had to create a model informed consent form (ICF) template. The process necessarily involved the consideration of major ethical issues of personal genome research, such as return of research results and incidental findings. First, we surveyed existing informed consent documents, and then came up with a draft of the ICF template based on a consideration of the Japanese context. To further refine the draft of the ICF template, we held repeated discussions with executive genome researchers of the MGSP. Through the discussion, we achieved some consensus regarding the research policy of the ICF template. However, for a key issue regarding research results and incidental findings—clear consent could not be achieved. On this issue, we confirmed that many genome researchers hold the following views: in Japan most personal genome studies aim toward an understanding of the cause and mechanism of diseases; and there is a great concern about the quality of research results for clinical purposes was usually beyond the scope of the original aims of research. As a general policy, we decided to leave the decision to each MGSP applicant. In this research activity, we crafted a research policy model that genome researchers in Japan, in which the cooperation between genome researchers in the MGSP was invaluable for incorporating pragmatic aspects of research. The lessons learned and experiences gained in establishing our research policy can be expected to share a number of similarities with those in other personal genome research and in clinical practice in Japan, and they may serve as a starting point for their policy-making activities.

2438T Mapping the second generation sequencing industry. M.A. Curnutt1, K.L. Frumovitz1, J. Bolinger1, G.H. Javit1, J., K.S. Carner2, D. Kaufman2, A.L. McGuire1. 1) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 2) Genetics & Public Policy Center, Washington, DC; 3) Sidney Austin, LLP, Washington, DC; 4) Berman Institute of Bioethics at Johns Hopkins, Baltimore, MD.

The next generation sequencing industry is rapidly evolving, as companies quickly transition from research services to clinical care. New business models are emerging, with companies offering a variety of services that have policy and regulatory implications. It is important to understand the development of the industry, and to anticipate policy issues that might impede clinical translation. We conducted a comprehensive web-based analysis with the aim of identifying the key business models within the next generation sequencing industry. This includes private companies offering a range of services from sequencing platforms to data analysis and storage.

We confirmed our findings with company representatives and experts in the fields of genetics and biotechnology to deepen our understanding of the industry. To date, we have focused on accurately capturing the companies’ customer base and practices and services. We present the main business models and discuss the regulatory implications of each. Our preliminary findings show that there is little competition amongst technology providers and increased market expansion in the areas of data analysis, storage, and informatics. As these technologies are incorporated into clinical diagnosis, we question how regulators will respond, and whether the modification of existing legal frameworks will be applicable or adequate. For example, will CLIA regulation apply to the complex analytics necessary to interpret sequence data, and if so are existing regulations appropriately tailored to these new methodologies? Will FDA assert jurisdiction over NGS-based laboratory testing, as it has over laboratory developed tests? We anticipate that NGS may enable the separation of the core phases of clinical laboratory testing—preanalytic, analytic, and postanalytic—into separate domains performed by different entities, which in turn may necessitate new models of regulation.

2437W Effect of Using a Family History Tool on Communication with Family and Health Care Providers. C. Wang1, A. Sen2, M. Pleague2, M. Ruffin3, S. O’Neill4, W. Rubinstein5, L. Acheson6, L. Kutny7 for the Family Healthware Impact Trial group. 1) Boston University School of Public Health, Boston, MA; 2) University of Michigan, Ann Arbor, MI; 3) Northwestern University, Chicago, IL; 4) National Institutes of Health, Bethesda, MD; 5) Case Western Reserve University, Cleveland, OH.

The family health history offers an ideal proxy to assess genomic risk and is the simplest applied genomic tool available. The Family HealthwareTM University, Cleveland, OH.

Trial group.

and Health Care Providers.

C. Wang

Effect of Using a Family History Tool on Communication with Family

and Health Care Providers.

S. O'Neill

J. Minari

Posters: Ethical, Legal, Social and Policy Issues in Genetics

350

A new large-scale project, the Genome Science Project, started in Japan in 2010. It is a five-year project funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The project aims to support various genome researchers in Japan who are funded by MEXT and selected through an annual open call; it does so by providing services of high-throughput DNA sequencing and high-grade information technology. A program of the Genome Science Project is the Genome Medical Group (GMSP), which collects human specimens and sequences the whole genome/exome to support the selected researchers, and thus naturally has to take into account various ethical, legal, and social implications. For this, a special group, Research Unit for Genome Ethics (GenEth), has been established within the project. In starting the MGSP, we had to create a model informed consent form (ICF) template. The process necessarily involved the consideration of major ethical issues of personal genome research, such as return of research results and incidental findings. First, we surveyed existing informed consent documents, and then came up with a draft of the ICF template based on a consideration of the Japanese context. To further refine the draft of the ICF template, we held repeated discussions with executive genome researchers of the MGSP. Through the discussion, we achieved some consensus regarding the research policy of the ICF template. However, for a key issue regarding research results and incidental findings—clear consent could not be achieved. On this issue, we confirmed that many genome researchers hold the following views: in Japan most personal genome studies aim toward an understanding of the cause and mechanism of diseases; and there is a great concern about the quality of research results for clinical purposes was usually beyond the scope of the original aims of research. As a general policy, we decided to leave the decision to each MGSP applicant. In this research activity, we crafted a research policy model that genome researchers in Japan, in which the cooperation between genome researchers in the MGSP was invaluable for incorporating pragmatic aspects of research. The lessons learned and experiences gained in establishing our research policy can be expected to share a number of similarities with those in other personal genome research and in clinical practice in Japan, and they may serve as a starting point for their policy-making activities.

2439W The commercialization of non-invasive prenatal testing: will a private market drive effective clinical translation? A. Aywah1, L. Sayres1, M. Cho2, R. Cook-Deegan2, S. Chandrasekharan3. 1) Duke University School of Medicine, Durham, NC; 2) Stanford University Center for Biomedical Ethics, Stanford, CA; 3) Duke University Institute for Genome Sciences and Policy, Durham, NC.

Since its 2001 US market introduction, non-invasive prenatal testing (NIPT) using cell-free fetal DNA found in maternal blood has provided the ability to analyze fetal genomes for genetic conditions without relying on commonly used invasive procedures. Unlike other prenatal tests, NIPT is being primarily driven by private industry, setting precedent for a new market landscape. The prenatal testing field, estimated to be worth up to 1.3 billion US dollars per year, provides strong commercialization incentive as a rapidly growing sector. Additionally, four US-based companies are offering NIPT for chromosomal aneuploidies as early as ten weeks into pregnancy. "NIPT on demand" is being marketed to pregnant women, and many of these have been exclusively licensed. All four are currently involved in patent infringement lawsuits, and interference proceedings are ongoing. The effect on availability and future NIPT development will depend on the resolution of these lawsuits, validation and scope of claims, and licensing strategies. It is possible that litigation related costs could increase prices, decreasing the cost-effectiveness of using NIPT and limit patient access. Also, companies may withhold collected data to secure a market advantage, potentially undermining quality assurance initiatives and decision making. Additionally, specific information about test utility in specific clinical contexts is not freely available. On the other hand, patents may help secure investment to drive future innovation in NIPT technology. While long-term effects are challenging to predict with ongoing litigation, it is necessary for physicians to better understand the implications of these cases. We present the main business models and discuss the regulatory implications of each. Our preliminary findings show that there is little competition amongst technology providers and increased market expansion in the areas of data analysis, storage, and informatics. As these technologies are incorporated into clinical diagnosis, we question how regulators will respond, and whether the modification of existing legal frameworks will be applicable or adequate. For example, will CLIA regulation apply to the complex analytics necessary to interpret sequence data, and if so are existing regulations appropriately tailored to these new methodologies? Will FDA assert jurisdiction over NGS-based laboratory testing, as it has over laboratory developed tests? We anticipate that NGS may enable the separation of the core phases of clinical laboratory testing—preanalytic, analytic, and postanalytic—into separate domains performed by different entities, which in turn may necessitate new models of regulation.

2437W Effect of Using a Family History Tool on Communication with Family and Health Care Providers. C. Wang1, A. Sen2, M. Pleague3, M. Ruffin4, S. O’Neill5, W. Rubinstein6, L. Acheson7 for the Family Healthware Impact Trial group. 1) Boston University School of Public Health, Boston, MA; 2) University of Michigan, Ann Arbor, MI; 3) Northwestern University, Chicago, IL; 4) National Institutes of Health, Bethesda, MD; 5) Case Western Reserve University, Cleveland, OH.

The family health history offers an ideal proxy to assess genomic risk and is the simplest applied genomic tool available. The Family HealthwareTM University, Cleveland, OH.

Trial group.

and Health Care Providers.

C. Wang

Effect of Using a Family History Tool on Communication with Family

and Health Care Providers.

S. O'Neill

J. Minari

Posters: Ethical, Legal, Social and Policy Issues in Genetics
Continuity of care of patients with inherited Genetic disorders. H. Azimt, S. Ghavimi. 1) Psychogenome, Nepean, Ontario, Canada; 2) Shahid Beheshti Medical University, Tehran, Iran; 3) All Saints University School of Medicine, Dominica.

Objective: Continuity of care of patients with inherited Genetic disorders. Methods: Survey of patients and physicians was completed by phone, email and in the hospital. Patients answered a series of questions based on the continuity of care they received from their physicians. Physicians in the survey answered a questioner based on how they felt they contributed to the patients overall health by knowing the knowledge of genetics necessary in order to have the best continuity of care for the patients. Results: Of 100 patients surveyed, 34 (34%) with MENIB syndrome, 11 (11%) diagnosed with Diabetes Mellitus type 2 and 25 (25%) diagnosed with Hemochromatosis. 40 patients (40%) had mixed disorders and their opinion was coupled to make one group. Out of the physicians surveyed, 34 Endocrinologist agreed that having the knowledge of Genetic, and studying their patients diseases allowed them to have a better a relationship with the patients in order to deliver the best continuity of care for them. Primary care physicians surveyed all agreed that cross multi-care for patients in the continuity of care practiced allowed them to better treat and deal with their patients diseases. All the physicians agreed that being up-to-date with the current genetics advancement they can have the best continuity of care for their patients. 100 Physicians participated in this survey. Conclusions: Knowing more about genetic disorders of our patients, will help to better build a relationship with them. This relationship can help the patient’s better cope and live a more productive healthier life. Physicians working as primary care physicians need to be up to date with current genetics advancement, and they need to make sure that the patients they care and treat that they are aware of genetics basis of their diseases as per this survey indicated.

Communicating with biobank participants: preferences for receiving aggregate results and providing updates to researchers. J. Master, M. Mercier, A. Goldenberg, R. Moore, C. Eng, R. Sharp. 1, 2, 3, 1) Genomic Med Inst, Cleveland Clinic, Cleveland, OH; 2) Dept of Bioethics, Cleveland Clinic, Cleveland, OH; 3) Dept of Genetic and Genomic Sciences, Case Western Reserve Univ, Cleveland, OH; 4) Taussig Cancer Inst, Cleveland Clinic, Cleveland, OH; 5) Center for Genetic Research Ethics & Law, Case Western Reserve Univ, Cleveland, OH.

Biobanks collect biological materials and/or health information for ongoing and future biomedical research. Guidelines recommend return of individual research results under certain criteria but do not discuss aggregate results or study updates. Previous work has found that biobank participants desire updates, but preferences as to the extent or manner have yet to be explored. Thus, we surveyed participants in a long-standing protocol-driven research biobank. Eligible participants were drawn from an IRB-approved study of patients with personal/family history suggestive of Cowden syndrome, a poorly-recognized hereditary condition predisposing to breast, thyroid, and other cancers. Participants gave blood and saliva samples and had no other interactions with researchers. The biobank had 3,618 participants at time of sampling. Survey eligibility included enrollment ≥18 years and within the biobank’s first 5 years, normal PTEN analysis, and contiguous United States address. Multivariate logistic regression analyses were performed to identify predictors of participant interest in internet vs. non-internet-based communication modes and communication modes allowing for participant/researcher interaction vs. one-way communication. Independent variables for the regressions included demographic characteristics and attitudinal variables logically or theoretically associated with desired modes of receiving general research updates. Variables were narrowed by independent Pearson correlations by cutoff p<.2. Variables with p<.02 were considered predictors of preferences. Surveys were returned from 840/1267 (66%) eligible subjects. Most (97%) wanted to receive general updates about the study with 92% wanting updates at least yearly. Participants wished to receive updates via paper (66%) or emailed (62%) newsletter with 85% selecting one of these options. 31% were not comfortable with any e-communication mode, with older, less educated, and lower-income persons strongly preferring offline approaches (p<0.001). Most (93%) had no concerns about receiving updates. 97% were willing to provide updates to researchers by post (71%) or email (50%). Plans to communicate results and updates should be made during study development and participants should be able to select if and how to receive them during informed consent. Study demographics may optimize communication modes. Funders should monetarily support the infrastructure needed for these efforts.


Imprinting disorders (IDs) are a group of rare congenital diseases affecting growth, development and metabolism with a lifelong impact on patients’ quality of life. Despite their common underlying (epi)genetic aetiologies, IDs are usually studied separately by small research groups working in isolation, and the basic pathogenesis and long term clinical consequences of IDs remain largely unknown. Efforts to elucidate the aetiology of IDs are currently fragmented across Europe and standardisation of diagnostic and clinical management is lacking. This COST Action will, for the first time, draw together researchers of all eight known human IDs in an interdisciplinary pan-European and worldwide Network for Human Congenital IDs, working to advance understanding of the pathophysiology with the major aim of translating this knowledge to improvement of diagnostic and clinical management for the benefit of the patients and their families. The Action will harmonise a common ID classification system, develop guidelines for treatment through consensus, create standard operation procedures (SOPs) for diagnosis based on best current practice, coordinate databases held in different countries to make them compatible and useful as a springboard for collective research initiatives, identify new imprinting disorders and causative mechanisms through collaborative effort, educate researchers and stimulate translational exchange. The ID network consists of >30 groups from >15 countries. It will join forces and complement studies to reduce health care costs and increase the life quality of patients suffering from IDs. The Action has started in June 2013, but is still open for further partners.

The ‘eugenics’ program and public health genomics in China and their implications towards East Asia. K. Muto, B. Zhao, H. Hong. 1, 2, 3) Dept Pub Policy, IMS, Univ Tokyo, Tokyo, Japan; 2) Graduate School of Frontier Sciences, Univ Tokyo, Tokyo, Japan.

Background: China has a high population growth rate, and thus faces great challenges in terms of population growth, health care and economic development. The State launched a family-planning policy in 1980, to control the size and raise ‘the general quality of the population’. As part of this family-planning policy, a new ‘eugenics’ program was developed to ensure the health of mothers and infants, and to improve the quality of the newborn population. As a result of economic development and improvements in medical treatment, the Chinese infant mortality rate dropped from 32 to 12 out of every 1,000 births in one decade. The National Medium- and Long-term Program for Science and Technology Development (2006-2020) set as a priority the acquisition of key technologies to ensure that the country’s population remains below 1.5 billion, and its birth defect rate below 3%. Purposes: To facilitate further discussions how both the new bioethics concept and the old disability concept have changed in China, we have looked back and analyzed the first criticisms of the Western science community with regard to the Chinese eugenics program of the 1990s and examine the future impact in East Asia. Methods: Using literature review methods in this study, we collected topical statements from academic societies, government documents, law, and guidelines. We also reviewed articles identified in PubMed and CNKI databases. Results: Western geneticists have fiercely criticized the program as an ‘abuse of genetics’ and a ‘violation of human rights’ (Morton, 1998). In contrast, some argue the law would help reduce births of physically or mentally abnormal babies. To look back at the history of the eugenics movement and its ensuing legislation, the term ‘eugenics’ is defined differently in China than it is in other nations. Discussions: Japan and Korea import non-invasive prenatal testing (NIPT) from the US, even though the two countries have different views and attitudes toward it. At present, China does not provide NIPT services. Although the ‘eugenics’ program is defined differently in China than in other nations, China argues against the use of a concept that is balanced ethically and scientifically.
2444T
The attitudes of patients with cystic fibrosis and their parents towards direct-to-consumer genetic testing. S. Janssens1, C. Bost2, I. Matheu2, A. De Poepel1, P. Borry2, 1) Center for Medical Genetics, University Hospital Ghent, Ghent, Oost-Vlaanderen, Belgium; 2) Master Life Sciences and Medicine, University of Ghent, Belgium; 3) Center for Biomedical Ethics and Law, University of Leuven, Belgium.

Background: Knowledge of carrier status enables prospective parents to make informed reproductive decisions. Recent technological advances in molecular genetics facilitate large-scale population carrier screening. In this prospect reflection is necessary about the desirability to offer community-based (preconception) carrier screening in the healthcare system. At this moment, commercial companies have started to offer preconception carrier tests directly to consumers. Methods: As part of a survey developed to evaluate attitudes towards preconception carrier screening for cystic fibrosis (CF), attitudes towards Direct to Consumer testing (DTC) were assessed. The study population was recruited from a register of 157 patients with CF who consult at least once in 3 months at the pneumology department at the University Hospital of Ghent, where one of the eight Belgian reference centers for CF is located. All the parents of CF patients under 18 years and all CF patients aged 16 years and older who attended the clinic in the period between August 13 and December 11 2012 were asked to fill in a questionnaire. An approval from the local research ethics committee was obtained. Results: In total 134 questionnaires were distributed of which 112 were returned. Response rate was 86.7% for the parents and 79.7% for the CF patients aged 16 years and older. 78.3% of the patients with CF and 87.1% of the parents had never heard about testing for carrier status of genetic diseases through the internet. Only 21.7% of the patients and 12.9% of the parents were aware of this possibility. Of the total survey population 57.7% finds that people have the right to order such a test directly at a commercial company. However, 57.1% of respondents disagrees that someone should be able to obtain the results without medical supervision. Only 43.2% of the patients with CF and 9.4% of their parents would buy a carrier test for other conditions then CF through a DTC genetic testing company. Discussion and conclusion: Awareness about DTC genetic testing for carrier status is low. Although a majority supports the right of individuals to receive test results without a physician. Only a minority would consider buying a carrier test from a DTC genetic testing company.

2444W
Genetic causal beliefs of morbidity: associations with health behaviors and outcome beliefs of behavior changes during two decades in the general population. A.H. Haukkala1, N. Hankonen1, H. Konttinen1, M. Perola2, H. Kääriäinen2, V. Salomaa2, 1) Social Research, University of Helsinki, Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland.

Background: The role and meaning of genetic information has increased considerably during the last decades. We examined changes in causal beliefs for morbidity and their associations with health behaviors and control beliefs from 1982 to 2002. Methods: Five population-based risk factor surveys from 1982 to 2002, participants aged 25 to 64 years (n=37,503). Subjects were asked to choose the most important cause for morbidity from the list of ten alternatives. Outcome expectancies of health behaviour changes were assessed with 2 items. Health behaviors included self-reported physical inactivity, current smoking and obesity based on measured height and weight. Findings: Prevalence of genetic causal belief increased from 4% (1982) to 10% (1992) and remained at that level until 2002. Older age groups were more likely to choose genetic causes while there were no associations with gender or education. Lack of exercise and overweight increased while inappropriate diet and smoking diminished as causes for morbidity during the study period. Smokers, physically inactive and obese respondents chose less often physical inactivity, smoking or diet as causes for morbidity and more likely genetic causes for morbidity. Genetic causal belief group had more pessimistic outcome expectations of health behavior changes for morbidity and more likely genetic causes for morbidity. Genetic causal belief group had more pessimistic outcome expectations of health behavior changes for morbidity and more likely genetic causes for morbidity. Genetic causal belief group had more pessimistic outcome expectations of health behavior changes for morbidity and more likely genetic causes for morbidity. Genetic causal belief group had more pessimistic outcome expectations of health behavior changes for morbidity and more likely genetic causes for morbidity. Genetic causal belief group had more pessimistic outcome expectations of health behavior changes for morbidity and more likely genetic causes for morbidity.

2446T
A systematic approach to the development of evidence-based family history screening in pediatric primary care. E. Edelman1, B.K. Lin1, N. Mikal-Stevens1, L. Vasquez2, K. Hughes3, J. Scott1, 1) NCHPE, Lutherville, MD; 2) March of Dimes, White Plains, NY; 3) American Academy of Pediatrics, Elk Grove Village, Il; 4) HRSA, Rockville, MD; 5) Harvard Partners, Boston, MA.

Background: While family health history (FHH) is recognized as an important risk assessment and management tool for primary care pediatric providers (PCPs), numerous barriers to its integration have been documented, including limited time, provider knowledge and confidence, and lack of guidance on how to best integrate tools to support FHH screening into the healthcare system. Objective: To determine conditions for inclusion on a novel electronic FHH and clinical decision support (CDS) tool for PCPs. Methods: A multi-method approach was used to determine inclusion of conditions for a pediatric FHH tool. Semi-structured interviews focusing on FHH practices were conducted with diverse PCPs (4 pediatricians, 1 family medicine physician, 1 family medicine and 1 pediatric nurse practitioner) recruited from the AAP Genetics in Primary Care Institute. Interviews were analyzed using qualitative thematic analysis. The published literature regarding PCPs’ use of FHH and the validity and utility for pediatric and adult conditions was reviewed. Practice guidelines and pediatric FHH forms (13 from pediatric practices and internet searches) were analyzed to determine FHH screening practices. Data on conditions were recorded in a database and analyzed using descriptive statistics. Practice findings and interview themes were summarized to inform condition selection. Criteria adapted from Yoon et al. (2003) were used to determine inclusion and focused on the clinical utility of screening in the general pediatric population. Results: Seventy-six conditions for possible inclusion on the panel were identified. PCPs considered most important were conditions of very high significance such as cancer, genetic conditions as well as chronic and mental illness. Conclusions: This is the first report to propose an evidence-based approach to determine conditions for pediatric FHH screening. This study presents a model for tool development that can be applied to other institutional or national efforts to develop FHH or genetic CDS tools for diverse clinics. These efforts are increasingly relevant to geneticists as they are called to provide guidance on FHH and genomic data standards in electronic health records and health information technology systems. Next steps include development of risk assessment algorithms and clinical decision support and identification of solutions for interoperability with electronic health records.

2447W
An Assessment of Perceived Medical and Psychosocial Needs of Families that have Children Affected by Duchenne Muscular Dystrophy in Madurai, India. S. Kcrijwal1, K. Ormond1, V.S. Arun2, L. Stanislas2, C. Siskind3, 1) Stanford University, 450 Serra Mall, Stanford, CA 94305; 2) Muscular Dystrophy Foundation India, 26-B, 1st Street, State Bank Colony-II, Bypass Road, Madurai - 625 010, Tamil Nadu, India.

Duchenne muscular dystrophy (DMD) is a hereditary progressive neuro-muscular condition that affects approximately 1 in 3500 male children worldwide. The aim of this study was to assess the medical, psychosocial and social needs for children and their families. Medical, social and cultural barriers in developing countries are likely to pose additional, unique needs for families. In order to determine these needs, we recorded, transcribed and analyzed 20 interviews with a total of 29 family members of children affected with DMD in Madurai, India. Participants were recruited based on variation in socioeconomic status, literacy, age and number of affected children, and number of deceased children. The main themes that evolved from the interviews centered on issues related to medical care, access and mobility, emotional difficulties and social support. Some aspects of these issues were specific to affected individuals and their families in India, whereas others were also experienced - albeit to a lesser degree - by their counterparts in developed countries. The former consisted of major inadequacies in healthcare, medical communication, psychosocial management and financial support. The latter consisted of underdeveloped solutions to mobility-related difficulties, a lengthy diagnostic process, and a lack of awareness amongst the medical community and society at large. In summary, our study revealed numerous unmet needs for families affected with DMD in Madurai, India. Long-term solutions to unmet needs may include: 1) policy change to improve healthcare funding and mobility infrastructure, 2) development of ancillary services such as genetic counseling and social work, 3) spreading awareness of government and non-governmental community and society at large, and 4) cross-talk between advocacy groups and health professionals. Relatively immediate solutions may include 1) guidance of families towards appropriate medical resources by advocacy groups, 2) increased investment of time in fund-raising activities to provide relief from financial strain, and 3) collaborations between muscular dystrophy advocacy groups and organizations that are invested in finding solutions to mobility needs of the disability community at large.
incorporate incidental CNVs?

Should the current ACMG recommendations for reporting incidental findings in whole genome analysis be any different?

disorders. Reporting of incidental findings is not new to medicine, so should recessive disorders, cancer susceptibility and dominant adult onset chromosome aneuploides, genomic disorders with incomplete penetrance of variants of unknown clinical significance in the proband, we incidentally found 28 (0.8%) clinically relevant CNVs including mosaic or full chromosome aneuploides, [2] genomic disorders with incomplete penetrance, [3] carriers of X-linked Mendelian traits, [4] carriers of autosomal recessive disorders, [5] cancer susceptibility and [6] dominant adult onset disorders. Reporting of incidental findings is not new to medicine, so should reporting incidental findings in whole genome analysis be any different? Should the current ACMG recommendations for reporting incidental findings incorporate incidental CNVs?

2448T

Reporting of incidental copy number variation (CNV) detected by chromosomal microarray analysis in 'normal' parents/family members - A review of 3500 cases. A. Patel, S.W. Cheung, P. Stankiewicz, A. Breman, S.R. Lalani, J. Smith, C. Shaw, S. Peacock, A. Braxton, L. Ellis, P. Ward, J.R. Lupski, A. Beaudet, W. Bi. Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX.

When interrogating the human genome, it is sometimes the case that the most significant finding is unrelated to the referring diagnosis. Such incidental findings can pose challenges with regard to which results should be revealed, to whom and by whom. Such issues have begun to be addressed with the recent ACMG recommendations for reporting incidental findings in Clinical whole Exome and Genome sequencing. While some incidental findings are currently of uncertain clinical significance, others are medically actionable and reporting of these incidental copy number variations (CNVs) can have a significant impact on risk counseling for future pregnancies and other family members at risk, as well as enabling interventions to improve health outcomes. In the course of performing chromosomal microarray analysis (CMA) on 3500 apparently 'normal' parents/family members for interpretation of variants of unknown clinical significance in the proband, we incidentally found 28 (0.8%) clinically relevant CNVs including [1] mosaic or full chromosome aneuploides, [2] genomic disorders with incomplete penetrance, [3] carriers of X-linked Mendelian traits, [4] carriers of autosomal recessive disorders, [5] cancer susceptibility and [6] dominant adult onset disorders. Reporting of incidental findings is not new to medicine, so should reporting incidental findings in whole genome analysis be any different? Should the current ACMG recommendations for reporting incidental findings incorporate incidental CNVs?

2449W

The Hospital for Sick Children Genome Clinic: Developing and evaluating a pediatric model for individualized genomic medicine. M.S. Krivonos1,2,3,4,5, S. Bowden1,2,3,4, N. Monfared1,2,3,5, D. Merico1,4, D.J. Stavropoulos1,2,3,4, M. Girdea1,2,3, R. Hayeems1, T. Stockley1,2,3, M. Szego1,2,3, G.D. Bader1,2,3, R.D. Cohn1,2,3,5, C.R. Marshall1,6,7, R. Zlotnik Shaul1,10,14, M. Brudno1,4, C. Shuman1,3,4, P. Ray1,2,4,7, 1) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 3) Division of Clinical and Metabolic Genetics, Dept of Paediatrics, Hospital for Sick Children, Toronto, ON, Canada; 4) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 6) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 7) Department of Paediatric Laboratory Medicine, Hospital for Sick Children, Toronto, ON, Canada; 8) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 9) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON, Canada; 10) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 11) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 12) Centre for Clinical Ethics; Providence Healthcare, St. Joseph’s Health Centre and St. Michael’s Hospital; Toronto, ON, Canada; 13) The Donnelly Centre, University of Toronto, Toronto, ON, Canada; 14) Department of Bioethics, The Hospital for Sick Children, Toronto, ON, Canada.

The transformative potential of whole genome sequencing (WGS) lies in its use in individualized genomic medicine, where knowledge of an individual’s genomic variants guides health care decisions throughout life. However, there are multiple challenges: Many disease genes have yet to be discovered; currently, the clinical utility of WGS is largely unknown; the clinical utility and socioeconomic impact of identifying medically actionable genomic variants (MAVs) in children are unknown; and existing counselling models are inadequate for clinical genomics.

To address these and other knowledge gaps we have created a paediatric ‘Genome Clinic.’ This multi-disciplinary research platform has several distinctive features, including: a) The project treats an individual’s genome as a resource of genetic information to be repeatedly queried in order to manage overall health, rather than a one time test performed to diagnose a pre-existing disorder. As a result, MAVs are a major focus of study rather than an inconvenient burden. b) Focusing on children offers the greatest potential for increased quality of life years. c) Project scope is broad - from development of new bioinformatics tools and discovery of new genes to health policy inquiries, assessment of models of clinical care and the ethics of consent. d) Project design emphasizes an interactive partnership between patients, parents and health care researchers in discovering the medical and psychological risks and benefits of genomic medicine. e) The Genome Clinic offers WES/WGS-based clinical research projects a ‘safe harbor’ for managing incidental findings.

Piloting our bioinformatics pipeline using a curated list of 2000+ disease genes found 5-13 MAVs and 5-18 carrier variants/individual requiring manual evaluation. We are enrolling 50+ children/year who are under investigation for a genetic disorder, along with their parents. Participants are counselled regarding WGS and given options of learning about specific classes of adult-onset MAVs and carrier variants. Variants relevant to the primary disorder are returned by the referring physician, while pharmacogenomic variants, MAVs, and carrier variants are disclosed to participants by Genome Clinic geneticists and genetic counsellors. Our systematic approach mitigates potential risks and maximizes benefits of this new clinical care model.
Frequency of ACMG recommended 57 gene incidental findings from whole exome sequencing in a cohort of 47 adult individuals. J. Wynn, M.L. Cremona, J. Martinez, Y.H. Cheung, W.K. Chung. 1) Department of Pediatrics, Columbia University Medical Center; 2) Department of Medicine, Columbia University Medical Center.

The American College of Medical Genetics (ACMG) recently recommended mandatory return of secondary findings for 57 genes in patients receiving clinical whole exome and whole genome sequencing (WES/WGS). Before implementation, the frequency of mutations in these genes in an unselected cohort should be determined to appreciate the impact of this policy on clinical practice. We completed WES on 47 adult individuals and analyzed results for secondary findings in the 57 ACMG genes. Variants in Human Gene Mutation Database were identified, and any variants with a minor allele frequency > 2% in the 1000 Genomes and Exome Variant Server were removed. The literature for all remaining variants was reviewed and only Class I variants using ACMG criteria were classified as mutations.

Five participants had previously had negative genetic testing of the BRCA1 and BRCA2 genes. For the other 42 subjects, the indication for WES was unrelated to the conditions on the ACMG list. A published pathogenic mutation in one of the 57 genes was identified in 13% (6/47) of participants. We identified 3 (6.4%) participants with SCN5A mutations; 2 of the 3 participants had no personal or family history suggestive of Long QT or Brugada syndrome. One participant had a history of two relatives who passed away suddenly <50 years. We identified 2 participants (4.3%) with mutations in cardiomyopathy genes (MYBPC3 and MYH7). One participant had recently had an echocardiogram showing mild concentric left ventricular hypertrophy with an intraventricular septal wall thickness of 1.3 cm (normal 0.6-1.1). The other participant was asymptomatic and had never had an echocardiogram. One subject had a mutation in the MSH6 gene. She had never been diagnosed with cancer but her mother had a history of colon and uterine cancer.

We found a frequency of secondary findings in 13% of adult participants for conditions that are believed to be rare. Few participants had a personal or family history of the condition identified by the secondary finding. These findings suggest problems with the accuracy of mutation classifications in the literature and/or inflated estimates of the penetrance for these conditions in an unselected population. Our results indicate there will be a significant burden to the clinical community to return secondary findings for these 57 genes, and perhaps implementation of any policy to return secondary findings should await curation of the mutation databases and genetic literature.
2452T

An assessment of the prevalence of rare nonsynonymous variants within the genes listed in the ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. S.P. Strom1, H. Lee1, J.L. Deignan1, K. Das1, E. Vilain2,3, W.W. Grody1,2,3, S.F. Nelson1,2,3, 1) Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 2) Human Genetics, UCLA, Los Angeles, CA; 3) Pediatrics, UCLA, Los Angeles, CA.

BACKGROUND: The American College of Medical Genetics recently released guidelines which recommend the reporting of incidental findings (IF) when performing clinical exome or genome sequencing. Specifically, the ACMG has stated that laboratories performing this type of testing should report ‘known pathogenic’ and ‘expected pathogenic’ variants within a list of 57 genes, all but three of which act in an autosomal dominant fashion. In these guidelines, the ACMG has estimated that 1% of individuals will carry a known or expected pathogenic variant in any one of these 57 genes. However, as no clinical grade database currently exists to assess which variants in each of these genes are known to be pathogenic, variant interpretation falls to each individual laboratory. Although guidelines state ‘these recommendations should not be construed as an expectation that the laboratory comprehensively assess these genes for all variants,’ compliance without such analysis requires currently unavailable variant annotation.

METHODS: We analyzed prior clinical exome sequencing results at UCLA from 226 de-identified individuals (either index or parental cases) to identify all high quality (Q>50) potentially pathogenic variants in any of the 57 IF genes. We define potentially pathogenic as: nonsynonymous variants with a population allele frequency <1%. RESULTS AND CONCLUSIONS: Contrary to the a priori estimate in the ACMG guidelines, we identified at least one potentially pathogenic variant within an IF gene in 194 out of 226 individuals (81%). 36% of variants identified are located at an HGMD annotated locus. A total of 294 unique variant loci, or approximately 1.3 per individual on average, were identified. As the combined disease prevalence of all IF gene disorders is far less than 81%, the majority of these variants are likely not clinically significant. However, identifying the subset of variants that are potentially significant represents an additional interpretive burden for the laboratory. Until clinical grade variant databases exist for all of these 57 genes, laboratories may need to report all potentially pathogenic variants in the IF genes rather than only reporting a select few, if the ‘fiduciary duty to prevent harm’ aspired in the guidelines is to be met. For now, laboratories attempting to implement these guidelines should work with ordering physicians to coordinate an appropriate approach for discussing incidental findings with each patient or family.

2453W

Do preferences matter? Creating and assessing a novel preference-setting tool for the return of genomic research results. P.L. Bacon1, S.K. Savage1, S.I. Ziniel2,3, K.D. Christensen2, N.L. Huntington1,2, E.R. Weitzman3,4, K.L. Taylor3,4, R.C. Green2,4, I.A. Holm1,2,3, 1) Program in Genomics, Boston Children’s Hospital, Boston, MA; 2) Department of Pediatrics, Harvard Medical School, Boston, MA; 3) The Manton Center for Orphan Disease Research, Boston Children’s Hospital, Boston, MA; 4) Children’s Hospital Informatics Program, Boston Children’s Hospital, Boston, MA; 5) Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, United States; 6) Partners Center for Personalized Genetic Medicine, Boston, MA; 7) Division of Developmental Medicine, Boston Children’s Hospital, Boston, MA; 8) Division of Adolescent Medicine, Boston Children’s Hospital, Boston, MA; 9) Program for Patient Safety and Quality, Boston Children’s Hospital, Boston, MA.

Whether and how research participants receive individual research results (IRRs) from genomic studies has generated great concern, and many argue that participants should be allowed to designate their preferences for IRR return. The goal of this study was to develop an online preference-setting tool for the return of IRR and to determine if parents anticipate greater benefit and satisfaction from enrolling their child in biobanks where they could set preferences for IRR versus biobanks where they could not. The preference setting tool was informed by a large survey conducted at Boston Children’s Hospital (BCH) assessing parents’ interest in enrolling themselves and their children in a hypothetical genetic research repository and preferences for return of IRR. The tool was further refined through cognitive interviews with parents of BCH patients to understand what dimensions of a condition mattered most with parents as they decided if they were interested in receiving a genetic result. The resulting preference setting model is based on a) the preventability of the condition and b) the severity of the condition. Parents may also opt-out of any of four categories of results: mental illness/psychological conditions; developmental disorders/learning disabilities; childhood-onset degenerative conditions; and adult-onset conditions for which there are no interventions during childhood. We then developed an online survey in which participants were randomized to be enrolled into one of four hypothetical biobanks with different results return policies: 1) participant receives no IRR; 2) participant receives all IRR; 3) participant chooses to receive all or no IRRs (‘binary choice’); and 4) participant sets preferences for IRR using the online tool (‘granular choice’). We pilot tested the survey with parents of children seen at BCH. Preliminary data suggest that, given a choice, the majority of parents elect to receive all possible IRRs about their children. Those parents who chose to receive all IRRs using the preference-setting tool rated the highest satisfaction of the 4 groups. In addition, the ‘granular choice’ group had the greatest proportion of parents stating that they would participate in a biobank that resembled the one they were randomized to. Preliminary findings will be confirmed in a forthcoming survey of 20,000 parents of patients at BCH.
Biobank Participants’ Perspectives on Aggregate Result Return. E. Bane1, E. Ludman2, J. Richards3, G. Jarvik4, S.M. Fullerton4. 1) Institute for Public Health Genetics, University of Washington, Seattle, WA; 2) Group Health Research Institute, Seattle, WA; 3) Medical Genetics, University of Washington, Seattle, WA; 4) Bioethics and Humanities, University of Washington, Seattle, WA.

Introduction: Researchers and policy makers have recommended returning, where feasible, aggregate (or summary) results of research conducted using biospecimens to biobank participants (Beskow et al. 2012). There is, however, lack of consensus on best practices for aggregate return, particularly about how to communicate participants’ perspectives on issues such as return. Here, we report a preliminary exploration of participant preferences and perspectives on the necessity for, and mode of, returning aggregate results of genomic biorepository research. Methods: Fifteen semi-structured interviews were conducted with participants in the Northwest Institute of Genetic Medicine (NWIGM) biorepository at the Group Health Cooperative and University of Washington in Seattle, Washington. NWIGM participants were interviewed by telephone about their perspectives on, and preferences for, return of results from research using human tissues. Interviews were recorded and de-identified transcripts subjected to a directed content analysis. Results: The participants interviewed reported a desire for individual results, and expressed a clear preference for allocating funds designated for return of summary results to additional research. In the event that aggregate results were to be returned, these participants preferred to receive them in the form of emails, letters, or websites rather than in-person interviews or phone calls. Conclusions: The results of this preliminary research indicate that while participants of genomic biorepository research are interested in research result participants view individual results as more relevant than aggregate results, and are more invested in the objectives and fruition of the research itself than in receiving results per se. These results suggest individuals should be consulted regarding preferences prior to formation of overarching policy regarding aggregate return, and indicate that further research with larger and more representative cohorts is warranted.

Should Secondary Findings From Whole Exome And Whole Genome Sequencing Be Released To Research Subjects? Our Ethical Responsibility. C.A. Campbell1, T. Bair2, D. Kolte3, P.J.H. Smith1,2. 1) Iowa Institute of Human Genetics, Iowa City, IA, USA; 2) Department of Otolaryngology - Head and Neck Surgery, University of Iowa, Iowa City, IA, USA.

Whole exome (WES) and whole genome sequencing (WGS) are powerful tools that have facilitated the identification of multiple genetic causes of human disease. Genetic variants identified incidentally to this primary discovery process are referred to as secondary findings. The American College of Medical Genetics (ACMG) has recently recommended that when WES and/or WGS are requested for clinical diagnostic purposes, the final variant report should include secondary findings in a defined set of 57 disease genes. These ‘Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing’ target genes primarily associated with hereditary cancer and cardiovascular disease since therapeutic intervention and/or screening make disease treatment and/or prevention possible. Analysis of data from the NHLBI Exome Sequencing Project on the Exome Variant Server (EVS) shows that there are 8,483 variants in European-American exomes in these genes. We will present the likelihood of identifying an ACMG-reportable variant in a research exome. As researchers, we must address this risk and the ACMG guidelines if we are to enroll patients with appropriate informed consent in genetic studies in which WES and/or WGS is planned. As the detection of variants depends on data quality, we present a summary of data quality for these genes using the SureSelectXT Human All Exon V4 (Agilent) capture and sequencing performed using paired-end chemistry on the HiSeq 2000 platform (Illumina, San Diego, CA) will be presented.

In the rapidly changing landscape of human genetics researchers are confronted with the dilemmas of: if, how, and when to give back research results to study participants. We will discuss the unique issues surrounding implementation of these guidelines in a research setting. In addition, the role of the genetic counselor in a research laboratory performing WES and/or WGS will be described. We propose select language in the informed consent document to deal with varying reporting options.
Human genetic researchers and biobank leaders support the return of high-risk, actionable research findings but face numerous impediments to responsible return of results. R. Dvoskin, J. Bollinger, K. Kregel, A.A. Padon, K.L. Edwards, D. Kaufman. 1) Genetics and Public Policy Center, Berman Institute of Bioethics, Johns Hopkins University, Washington, DC; 2) University of Washington, School of Public Health, Department of Epidemiology, Seattle, WA.

**Background:** Biobanks and genetic cohort studies are important tools in the search for genetic and environmental factors underlying a broad range of diseases. The data generated by these resources may include genetic information of potential significance to study participants. Whether and which, individual research results should be returned are under heated debate. As policies and recommendations evolve, it is useful to consider the opinions and practices of the research and biobank communities.

**Methods:** 59 U.S.-based genetic researchers and biobank leaders were interviewed about human subjects issues including the return of research results. Semi-structured interviews were transcribed, coded, and analyzed for themes. Reasons for and against returning results were summarized.

**Results:** A majority of interviewees were not opposed in principle to the idea of returning some medically useful research results. Interviewees raised a number of arguments against their return, however, most of which fell into five categories: practical and logistical reasons; the uncertain quality and meaning of the data; excessive cost and personnel burdens; the risk of doing more harm than good; and the lack of guidance regarding research and biobank obligations. Two common reasons for returning results were a perceived duty to disclose information that could benefit someone's health, and participants' right to their own information. Additional reasons included participation in research, an interest in the results of individual research, the interests of the research participants and whether participants are clinical patients or healthy volunteers.

**Conclusions:** Concerns about uncertain or unconfirmed results and ambiguity about researchers' obligations may influence decisions to return results. In addition to providing guidance on what results to return, future research should consider the importance of contextual factors and address the practical, financial, and ethical challenges to disclosing research results. Researchers' support for the return only of well-understood genetic findings and for transparency about the quality and meaning of research data agree with public preferences observed in some previous studies.
User satisfaction with a web-based tool for self-guided management of results from ES/WGS. 

**2462T User satisfaction with a web-based tool for self-guided management of results from ES/WGS.**  

A major obstacle to taking full advantage of exome and genome sequencing (ES/WGS) data in a clinical setting is how to effectively manage incidental or secondary results. Use of traditional approaches (e.g., face-to-face interview with a provider) for each result that could be offered for return is simply untenable. Over the past several years, we have developed a return of result strategy based on self-guided management of results using a web-based tool called My46 (http://www.my46.org). My46 enables users to set preferences for which results they want to receive, to change their preferences over time, to review standardized information about each trait for which a result is available, and to receive each result privately and conveniently. The usability of such a tool partly depends on whether individuals can successfully complete each task needed to manage results return and whether they are comfortable doing so. We tested the usability and satisfaction of setting ES/WGS result preferences with My46 in adults (n=47) with Mendelian or complex diseases enrolled in ES projects. Participants were primarily European American (94%), female (83%), and 26 to 81 years old (mean=44). The mean composite scores for a modified Computer System Usability Questionnaire were highly positive, on a scale of 1 to 5 with 5 being most favorable: usefulness=4.68 (3.68-5.00), information quality=4.58 (3.14-5.00) and interface quality=4.46 (3.00-5.00), with an overall score of 4.61 (3.68-5.00). The majority of participants (96%) indicated it was simple to use My46, 91% found the information provided in My46 easy to understand, and 81% indicated My46 had all the functions and capabilities they expected it to have. Most participants (75%) indicated they were satisfied with the use of My46 as a way to select preferences, thought My46 had the right kind of information to aid in decision making, and would recommend My46 to someone else for managing genetic results. A subset of individuals (n=21) were interviewed after setting their preferences and all reported that the site was easy to use, simple, thorough and user-friendly. These individuals valued that setting their preferences using My46 was fast and efficient, though some expressed concern that use of confirmation ‘pop-ups’ was unnecessary. These results suggest that usability and satisfaction with My46 are high, and underscore the need for further development and testing of web-based approaches for ES/WGS results management.

**2463W The Industry Pharmacogenomics Working Group (I-PWG) Perspective on Providing Individual Research Results and Incidental Findings to Clinical Trial Research Participants.**  
S.K. Prucka, J.L. Arnold, J.E. Brandt, S. Gilardi, L.C. Harty, F. Hong, J.S. Malia, D.J. Pulford

The pharmaceutical industry has not been a prominent figure in recent discussions regarding the return of individual genetic research results and incidental findings. The Industry Pharmacogenomics Working Group (I-PWG) however, is well-positioned to comment on issues related to providing genetic research results that are generated from clinical trials. The I-PWG consists of representatives from 20 pharmaceutical companies engaged in pharmacogenomics research as part of product development. Industry-sponsored genetic research is typically directed at understanding the safety and efficacy of the therapeutic compound under development and researching the genetics of disease to better define subgroups that are the most likely to receive a benefit. This research progresses alongside the development of a therapeutic compound from pre-clinical work through clinical trials and often continues with post-marketing studies after product is launched on the market. The majority of this research is exploratory in nature and many years may pass before the clinical implications can be confirmed and validated. Without clear clinical relevance, the benefit of providing genetic research results to participants is unclear. Further, given the exploratory nature of this research, many of the assays employed may not yet have been fully validated or meet the local requirements for sharing with research participants. When results are provided, the Investigator at each clinical trial site is responsible for delivering this information. Investigators’ varying levels of genetic literacy can make it challenging to communicate the information in a side-effect information sheet or report, and industry has not been a prominent figure in recent discussions regarding the return of individual genetic research results and incidental findings.
Interest in different types of individual genome sequencing results among younger breast cancer patients. M. R. Ray1, J. Ivanovich1, B. B. Biesecker2, L. G. Dressler3, R. Dresser4, M. S. Goodman5, P. J. Goodfellow6, K. A. Kaphingst7, 1) Division of Public Health Sciences, Washington University School of Medicine, Saint Louis, MO; 2) Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD; 3) Mission HealthCare, Asheville, NC; 4) Washington University School of Law, Saint Louis, MO; 5) College of Medicine, Ohio State University, Columbus, OH.

Background: Critical communication challenges arise if individual results from genome sequencing are returned. Patients diagnosed with breast cancer at a young age are a key population for early application of genome sequencing to identify susceptibility alleles and variants affecting treatment response. Empirical data are needed regarding what genome sequencing results are of interest to these patients to inform ongoing policy discussions for return of results. Methods: We conducted 48 semi-structured individual interviews with women diagnosed with breast cancer at age 40 or younger to investigate their interest in different types of genome sequencing results. We stratified recruitment by family history of breast cancer and BRCA1/2 mutation status to examine differences across subgroups. Interviews focused on interest in return of individual results for six types of gene variants: affect risk for a preventable disease; affect treatment response; uncertain clinical significance; affect risk for a disease that cannot be prevented or treated; carrier status; and ancestry or physical traits. Descriptive statistics were used to examine types of variants of greatest and least interest. Qualitative thematic analysis of interview transcripts was conducted using NVivo. Results: Participants identified variants that affect risk for a preventable disease (81%) or affect treatment response (17%) as those of greatest interest, often due to clinical actionability: ‘...you can delay the onset of any real symptoms or issues...it’s always gonna come back to quality of life.’ Participants had the least interest in variants without a health meaning (56%) or that affect risk for an unpreventable or untreatable disease (27%). Lack of interest in the latter type often related to quality of life: ‘If there’s nothing I can do about it, I’d rather live as happy a life as possible.’ Participants differed in their interest in variants of uncertain clinical significance, with some suggesting the information might have meaning in the future, while others thought it could cause unnecessary worry. Conclusions: Participants had the greatest interest in clinically actionable results; results that might negatively impact quality of life were of less interest. However, interest in different types of results, and perceptions of clinical and personal utility, varied. Policies for return of individual genome sequencing results should take into account differences in patient preferences.
2466T
Evaluation of ACMG Recommended Incidental Findings in Clinical Whole Exome Sequencing. E. Haferfield1, A. Daly2, A. Fuller1; N. Smaouli3, K. Retterer1, P. Vitazka1, G. Richard1, W. Chung2, S. Bale1; 1) GeneDx, Inc., Gaithersburg, MD; 2) Columbia University Medical Center, New York, NY.

Whole exome sequencing (WES) has the potential to identify incidental findings, or genetic variants that can result in medical implications unrelated to the primary indication for testing. The American College of Medical Genetics and Genomics (ACMG) recently recommended reporting incidental findings in 57 genes for individuals undergoing WES. These 57 genes are primarily associated with an increased risk for cancer or cardiac disease. The ACMG estimated that ~ 1% of patients undergoing WES will have reportable incidental findings. To evaluate this, the frequency of incidental findings in these 57 genes was examined in 50 de-identified probands who had WES and did not have cardiac disease or cancer as a test indication. The 57 genes were assessed for sequence changes present at ≥ 2% frequency in the 1000 Genomes database, and further reviewed for variant frequency in sub-populations, evidence for pathogenicity in the literature, and consistency with the expected mutation spectrum of the gene. The initial screen yielded 345 variants, of which 318 variants were eliminated based on high frequency in at least one ethnic sub-population, leaving 27 variants requiring further assessment. Of those, ten mutations (one in each of 10 individuals) met the criteria outlined by the ACMG to be reportable as known pathogenic (KP) or expected pathogenic (EP), demonstrating that 20% of WES cases have a reportable incidental finding. Of the 10 mutations, two were in cancer-related genes, six in cardiac-related genes and one mutation each was identified in a gene associated with hypercholesterolemia and malignant hyperthermia. Identification of incidental findings in 20% of patients undergoing WES has significant implications for ordering clinicians, genetic counselors and the clinical laboratories providing testing. Clinicians who return incidental findings to probands and families have an increased responsibility for pre and post-test counseling as well as medical follow-up, while clinical laboratories must integrate additional bioinformatic and analytic resources into these evaluations. The costs associated with the additional efforts could be substantial. Going forward, it will be important to continue to evaluate the impact of reporting incidental findings on patients and providers.

2467V
Giving and receiving: comparing parents’, paediatricians’ and genetic health professionals’ opinions about uncertain chromosomal microarray results. E. Turbitt1,2, D. Amor1, J. Halliday1, S. Patil1, M. Massie1, R. Rau-Murphy1, M. Cornes1, C. Manschreck1, J. Vijai6, Z. Stadler1, S. Lipkin7, K. Offit1,5,6, M. Robson1; 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY; 3) University of Toronto, Li Ka Shing Knowledge Institute, St. Michael’s Hospital, Toronto, Ontario, Canada; 4) Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, NY; 5) Weill Cornell Medical College, New York NY; 6) Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, NY.

Background: Conceptual frameworks have been developed to guide the disclosure of incidental results from whole genome or exome sequencing (WG/ES), however limited tools exist to support patient decision-making about the receipt of large volumes of incidental data. We aimed to: (1) develop and evaluate a decision-aid categorizing incidental WG/ES results and (2) explore perceived utility and decision-making regarding the categorization of results. Methods: We developed a decision-aid based on Berg et al.’s ‘binning’ scheme. Variants are categorized by their level of actionability and potential for distress. Bin 1 includes variants with established disease prevention or treatment guidelines (as the ACMG recent guidelines); Bin 2 includes clinically valid variants without established interventions and is further divided into 2a,b,c (refer to companion abstract by Schrader for binning methodology); Bin 3 includes carrier results; Bin 4 contains variants of unknown significance. We are pre-testing the decision-aid through interactions with 40 parents recruited from hospital waiting rooms. The decisional conflict scale assesses satisfaction and effective decision-making. Data analysis includes qualitative content analysis and descriptive statistics. Results: Preliminary results demonstrate wide variation in selection and perceived utility of the bins. Participants favored Bin 1 because results were actionable. They perceived Bin 2a unlikely to have major health/lifestyle impact. Bin 2b was difficult to understand because it consists of a heterogeneous group of variants (carrier, diagnostic and progressive) with varying potential for distress. Participants perceived Bin 2c as more distressing, precluding selection of Bin 2b. Selection of Bin 2a and 2c was driven by motivations to employ risk-reducing behaviors and inform life plans. Selection of Bin 3 varied by attitudes towards reproductive interventions and potential for distress. Bin 3a,b,c were perceived as more distressing, precluding selection of Bin 2b. Selection of Bin 2a and 2c was driven by motivations to employ risk-reducing behaviors and inform life plans. Selection of Bin 3 varied by attitudes towards reproductive interventions and potential for distress. Participants had mixed views about shared results with relatives, doctors and among researchers. Conclusion: Preliminary results suggest that this decision-aid can foster informed decision-making, which involved doctors and among researchers. Future work will evaluate decisional conflict and potential re-conceptualization of Bin 2.
2469W

Actionable incidental findings in the 1000 Genomes dataset. E. Olsson1, C.E. Cottrell2, N.O. Davidson2, N.O. Stitzel3,5, L. Chen4, S. Hartz5, S. Kouf6, R. Nagarajan7,6, N.L. Saccone5, L.J. Bierut1. 1) Department of Psychiatry, Washington University School of Medicine, St Louis, MO; 2) Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO; 3) Division of Gastroenterology, Department of Medicine, Washington University School of Medicine, St Louis, MO; 4) Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St Louis, MO; 5) Division of Statistical Genomics, Washington University School of Medicine, St Louis, MO; 6) Department of Genetics, Washington University School of Medicine, St Louis, MO.

Purpose: Advances in genomic sequencing have increased the likelihood of uncovering incidental findings. The goal of this study was to provide a conservative estimate of the frequency of actionable incidental findings using the 1000 Genomes dataset and to present a multi-stage systematic approach for identifying these findings. Our approach not only focused on diseases that several specialists agree should be returned as incidental findings in clinical sequencing studies, but also incorporated literature review and expert opinion to provide a lower bound estimate of the prevalence of incidental findings. Methods: Candidate single nucleotide variants in 17 genes associated with 11 dominant actionable conditions were extracted from the 1000 Genomes Browser. These variants were filtered using the Human Gene Mutation Database and defined parameters, appraised through literature review, and examined by a clinical laboratory specialist and expert physician. Only variants that the clinical specialists agreed were pathogenic were identified as incidental findings. Results: From the 1000 Genomes Browser, 22,182 single nucleotide variants were extracted from the 17 genes, and filtering led to 119 candidate variants. Literature review refined this number to 11 variants, and clinical specialists agreed that 2 were likely pathogenic: LDLR p.Trp4+ associated with familial hypercholesterolemia and KCNH2 p.Leu552Ser associated with Long QT syndrome. For 3 variants, 2 in APC and 1 in MLH1, there was discordance between the pathogenicity predictions of the specialists: while the clinical laboratory specialist determined that they were variants of unknown significance or rare polymorphic variants, the expert physician identified these variants as probably pathogenic. Conclusion: Candidate single nucleotide variants in 17 genes associated with 11 dominant actionable conditions were extracted from the 1000 Genomes Browser. These variants were filtered using the Human Gene Mutation Database and defined parameters, appraised through literature review, and examined by a clinical laboratory specialist and expert physician. Only variants that the clinical specialists agreed were pathogenic were identified as incidental findings. From the 1000 Genomes Browser, 22,182 single nucleotide variants were extracted from the 17 genes, and filtering led to 119 candidate variants. Literature review refined this number to 11 variants, and clinical specialists agreed that 2 were likely pathogenic: LDLR p.Trp4+ associated with familial hypercholesterolemia and KCNH2 p.Leu552Ser associated with Long QT syndrome. For 3 variants, 2 in APC and 1 in MLH1, there was discordance between the pathogenicity predictions of the specialists: while the clinical laboratory specialist determined that they were variants of unknown significance or rare polymorphic variants, the expert physician identified these variants as probably pathogenic. From the 1000 Genomes Browser, 22,182 single nucleotide variants were extracted from the 17 genes, and filtering led to 119 candidate variants. Literature review refined this number to 11 variants, and clinical specialists agreed that 2 were likely pathogenic: LDLR p.Trp4+ associated with familial hypercholesterolemia and KCNH2 p.Leu552Ser associated with Long QT syndrome. For 3 variants, 2 in APC and 1 in MLH1, there was discordance between the pathogenicity predictions of the specialists: while the clinical laboratory specialist determined that they were variants of unknown significance or rare polymorphic variants, the expert physician identified these variants as probably pathogenic.

2470T

Binning framework for a decision aid for the selection of incidental genome sequencing results. K. Schrader1, Y. Bombard1,2,4, E. Glogowski1, M. Salerno1, S. Palet2, M. Massie5, R. Rau-Murthy1, M. Corines1, C. Manshreck1, J. Vijai1, Z. Stadler1, L. Lipkin1, K. Offit1,2,6, M. Robson1, (first and second authors have equal contributions). 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, NY; 3) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY; 4) University of Toronto, Li Ka Shing Knowledge Institute, St. Michael’s Hospital, Toronto, ON; 5) Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, NY; 6) Weill Cornell Medical College, New York, NY.

Trepidation of disclosure of incidental genomic results stems from a perceived difficulty in adequately counseling patients about the broad scope of available findings and implications. To simplify this process, Berg [2011,2013] pioneered the concept of ‘binning’ incidental results based on medical actionability (Bin1) or, in its absence, the potential for psychological harm with disclosure (Bin2a-c). We have adapted these bins and propose further modifications to attempt to improve patient understanding and decision-making. Berg’s bins acted as a guiding framework for our decision aid, and were adapted by consensus of our working group of medical and cancer geneticists, genetic counselors, and a public health genomics researcher. Adaptations include: adding Bin3, which allows patients to learn healthy carrier status (for diseases on New York State’s newborn screening panel) without the rest of Bin2b; moving genes identified as medically actionable by the ACMG [2013] to Bin1, recommending but not mandating Bin1 disclosure; limiting Bin2a SNPs to those used by Coriell Personalized Medicine Collaborative, limiting the number of pharmacogenomic SNPs, and limiting Bin2c to few progressive diseases with which the lay population is generally familiar. Compared to Berg’s bins, Bin2b still contains disease genes ineligible for the other bins, and Bin2c contains selected genes with the highest subjective risk for psychological harm. Bin4 still contains uninterpretable data not available for disclosure. Determination of medically actionable genes is an iterative process and concerted efforts have developed frameworks for doing this in a reproducible manner. Determining which genes may cause psychological harm and comprise divisions of Bin2 is decidedly more subjective. Adapted bins are pretested in the decision aid (refer to companion abstract Bombard et al). Preliminary findings suggest patients have difficulty accepting the broad implications of Bin2b, thus we propose remodelling using an alternate framework to separate genes into three inherent divisions 1) Diagnostic: suggesting static, congenital conditions that are potentially diagnosable with further clinical investigation 2) Carrier: with no anticipated phenotype and 3) Predictive: suggesting risk for future disease eg. progressive degenerative disorders. Patient data will guide whether additional parsing (or merging with other bins) may enable patients to make more informed decisions.
How Interested are Parents in Genome Screening for Their Newborns?

R.C. Green1,2, S.S. Kalia3, H.L. Levy4, I.A. Holm5, S.E. Waisbren1,3, 5, 6, S.E. Karlawish2, T.O. Obisesan4, L.B. Waterston5, L.A. Cuples3, W.R. Uhlmann5, E. McCarty Wood7, R.C. Green1,8 on behalf of the REVEAL Study Group, 1) Dept Medicine, Div Genetics, Brigham & Women's Hosp/Harvard Med, Boston, MA; 2) Department of Health Behavior & Health Education, University of Michigan School of Public Health, Ann Arbor, MI; 3) Department of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Medicine, Howard University, Washington, DC; 5) Departments of Biostatistics & Epidemiology, Boston University School of Public Health, Boston, MA; 6) Departments of Internal Medicine & Human Genetics, University of Michigan, Ann Arbor, MI; 7) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 8) Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA.

Background: Parents of newborns are highly stressed while in the hospital after delivery, yet this would be the most efficient time to enroll such parents into empirical studies of newborn genome screening (NGS). Concerns about enrollment at this time include whether discussing this topic would cause parents to question or refuse state-mandated newborn screening (NBS), and whether parental interest in NGS research would be the same if queried several months later, after the post-partum period. Methods: We surveyed parents of healthy neonates on the post-partum unit at Brigham and Women’s Hospital about their interest in participating in research where their infant would receive NGS. Parents rated their interest on a 5-point Likert scale from ‘not at all interested’ to ‘extremely interested’. We are currently re-contacting these parents and have so far surveyed 66 parents at 3-6 months after delivery with the same questions. Results: Parents were systematically approached while on the post-partum unit and 87.3% agreed to participate. Among 216 parents queried about their interest in NGS, none questioned or declined state-mandated NBS. Responses were dichotomized to ‘not at all’ interested in NGS, versus ‘somewhat’, ‘very’, or ‘extremely’ interested. In the hospital, 182 (84.3%) parents were at least ‘somewhat’ interested in NGS, and 101 (46.8%) were ‘very’ or ‘extremely’ interested. Level of interest was not significantly associated with parents’ age, sex, race, education, income, or number of biological children. Among the 66 parents surveyed again after 3-6 months, 53 (80.3%) were at least ‘somewhat’ interested in NGS in the hospital vs. 55 (83.3%) at follow-up. However, the intensity of interest declined slightly over time; 32 (48.5%) were ‘very’ or ‘extremely’ interested in the hospital vs. 27 (40.9%) at follow-up. At 3-6 months, 58 parents (87.9%) remained equally interested in NGS, while 3 (4.5%) became less interested, and 5 (7.6%) became more interested. Conclusions: These preliminary data suggest that parents of healthy neonates can be surveyed about NGS in the hospital without threatening compliance on conventional NBS. Interest in research that would involve NGS is high among parents of newborns in the hospital, with at least half reporting that they would be ‘very’ or ‘extremely’ interested in such research. Interest in NGS appears to remain high 6 months later, though the intensity of the interest may decline slightly over time.
2473W The impact of genetic risk disclosure for Alzheimer's disease: Findings from the REVEAL Study APOE ε4 homozygotes. L.B. Waterston1, J.H. Karlawish2,3, C.A. Clark4, R.C. Green1, 1) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Division of Human Genetics, Department of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Medicine, Weill Cornell Medical College, New York University School of Public Health, Ann Arbor, MI; 4) Boston University School of Public Health, Boston, MA; 5) Partners Healthcare for Personalized Genetic Medicine, Boston, MA.

Background: It has been shown that individuals at increased risk for Alzheimer's disease (AD) because they carry at least one copy of the APOE ε4 allele experience greater test-related distress, perceive greater AD susceptibility, and make more changes to health behaviors than ε4 negative individuals. Differences are to date not well demonstrated among individuals homozygous for ε4 (ε4HM) and ε4 heterozygotes (ε4HT), however. Methods: Cognitively normal adults (mean age 57; 64% female; 84% White, 15% African American; 88% with family history of AD) across three randomised trials learned their APOE genotypes and AD risk estimates (range: 6–77%) based on their genotype, gender, ethnicity, and family history. Primary outcomes measured in 6 weeks, 6 months and 1 year post-disclosure were symptoms of anxiety, depression, and test-specific distress. Also assessed at one year were perceived AD risk (proportions endorsing 'high' or 'very high'), self-reported behavior changes (diet, exercise, taking medications or vitamins) and overall impact of risk disclosure. Results: Of 648 genotyped participants, 399 were ε4Neg (62%), 221 were ε4HT (34%), and 28 were ε4HM (4%). ε4Neg, ε4HT and ε4HM's did not differ in serial depression or anxiety measures. In survey 3, 31 were more likely than others to report test-related distress and ε4HT and ε4HM's at all time points (all p<0.001), but ε4HT and ε4HM's did not differ on this measure (all p>0.05). Differences in percentage of ε4HT and ε4HM's perceiving their AD risk as high were not evident (64% versus 72%; p>0.05). Comparing these activities and present science and social situations, it may come across as being a short period compared to the history of many scientific disciplines, the world of biology significantly changed with these new genetics and genomics. Furthermore, this new science gave the great impacts to bioethical and ELSI activities that appeared in that almost same period in Japan. We first overview a history of bioethical and ELSI activities focusing on the following four kinds of activities: the bioethics committees of the sciences, the ethics review boards in the institutes, the specialist groups or committees for ELSI in the large-scale research projects, and the bioethics in the academic sector. From these viewpoints, it was shown that the people or groups have created new ways or changed old ones to respond to various situations and present science and social situations, it was also revealed that there are a few remaining issues that have been considered less, therefore have not been fully introduced in the formal settings, such as the right of the populations, the right of the people other than a direct relationship to the donors of DNA, the long-term storage and reuse of the data, and the ethical status of human genome itself. In addition, we looked back the activities of another category of actors, that is, the scientific disciplines which they participate in an additional genetic counseling session and focus group. We will also complete the overall impact as positive (74%, p<0.001), but ε4Neg's were more likely than ε4HT's and ε4Neg's to report 12-month changes in diet (61% versus 34% and 27%), exercise (58% versus 30% and 28%) and taking medications or vitamins (58% versus 38% and 27%) (all p<0.001). ε4Neg's were more than other groups likely to have the overall impact as positive (74%, p<0.001), but ε4HT's and ε4HM's did not differ in their percentages rating the overall impact as positive (50% vs. 50%). Conclusions: Disclosure of ε4 homozygote genotype was not associated with increased perceived risk of AD, depression, anxiety, test-specific distress, or overall impact ratings compared to being a heterozygote; but it was correlated with greater likelihood of engaging in putative AD risk reducing behaviors. Such findings can inform future efforts to disclose APOE genotype to ε4 homozygotes in research trials, consumer settings or clinical practice.

2475W Results of the bioethics questionnaire survey of Nepali university students on genetic testings, H. NUMABE1,2, R. POKHAREL1 1) Department of Genetic Counseling, Graduate School of Humanities and Sciences, Ochanomizu University, Bunkyoku-ku, Tokyo, Japan; 2) Department of Paediatrics, Kyoto University Hospital, Kyoto, Japan; 3) Department of Orthopaedics, Institute of Medicine, Tribhuvan University Teaching Hospital, Kathmandu, Nepal.

In Asian countries such as Japan, Korea, and China, genetic testing has been carried out using similar methods with accuracy comparable to western countries. However, the implementation of testing is influenced by various factors such as population policy, economic situation, and the medical system of each government, as well as the bioethical view of each individual. Especially, culture plays a substantial role in genetic medicine and reproductive medicine. It is thus imperative to take cultural influence into consideration when giving genetic counseling and genetic testing to patients and clients in Japan and in the other Asian countries. We made the questionnaire survey of Japanese university students on genetic diagnoses, and reported the result at the ASHG Meeting in 2008. The questionnaire was consist of properties of the student, considerations to spiritual matters, acceptability of an artificial abortion, the time of the beginning of human life, and considera-


Background: Next-generation sequencing technologies have advanced rapidly over the past decade, resulting in a dramatic decrease in the cost of whole genome and whole exome sequencing. Although these technolo-

2474T A historical overview and a reflection on bioethical and ELSI activities for basic researches in human genetics and genomics in Japan. N. Yamanouchi, K. Nomura, K. Watanabe, Y. Iwasaki. National Institute of Genetics, Graduate School of Medicine, Osaka University, Suita, Osaka, Japan.

Over 20 years have passed since the human genome project started. Including genetic-recombination era, it is more than half a century. Though it may come across as being a short period compared to the history of many scientific disciplines, the world of biology significantly changed with these new genetics and genomics. Furthermore, this new science gave the great impacts to bioethical and ELSI activities that appeared in that almost same period in Japan. We first overview a history of bioethical and ELSI activities focusing on the following four kinds of activities: the bioethics committees of ministries, the ethics review boards in the institutes, the specialist groups or committees for ELSI in the large-scale research projects, and the bioethics in the academic sector. From these viewpoints, it was shown that the people or groups have created new ways or changed old ones to respond to the requirements conducting the research appropriately. For example, the fundamental rights of the research participants, such as the donor autonomy and privacy protection, became guaranteed in the written research guide-

nations to various genetic testing. We made the similar questionnaire survey in Nepal in 2012. We make an interim report. In Nepal, 70 medical students (65 Hindu, 2 atheists, 1 Buddhist, 1 Muslim, 1 Jewish person) answered our questionnaires. Fifty students recognized that the human life begins from the fertilization. Acceptances of genetic testing were as follows: Clinical genetic testing 91%, [Postnatal] Predictive Genetic Diagnosis 74%, Genetic Carrier detection 84%, Prenatal diagnosis 82%, Preimplantation genetic diagnosis (PGD) 81%. We also compare the results of Japanese and Nepali.
2477F Parental decisions following prenatal diagnosis of chromosomal abnormalities around Nagoya, Japan. N. Suzuki,1,2, K. Kuma1,1, S. Goto1,1, A. Nakamura1,1, S. Saito1,1, M. Sugiuira-Ogasawara1,1 1) Nagoya City University, Nagoya, Japan; 2) Chiba University of Commerce, Chiba, Japan.

Indications for amniocentesis at 15-18 weeks of gestation include advanced maternal age, maternal serum screening results, ultrasonographic findings for congenital malformations or markers of aneuploidy, and chromosomal abnormality in a previous pregnancy. Cytogenetic karyotype analysis by amniocentesis is highly accurate as compared to cell free fetal DNA testing, which has become clinically available in a few countries for women at increased risk of fetal aneuploidy. When an abnormal karyotype is identified by prenatal diagnosis, parents are faced with decision options that include termination versus continuation of the pregnancy. Decisions concerning prenatal testing and termination of pregnancy in case of affected fetuses are complex and may be influenced by a variety of factors, such as the country’s health system, its abortion laws, as well as social and cultural backgrounds. Parental decision-making to terminate or continue a pregnancy was studied after prenatal diagnosis of a chromosome aneuploidy among a sample of patients around Nagoya, Japan. The 1051 amniocentesis cases at 15-18 weeks of gestation were analyzed. The 60 cases of chromosomal anomalies with aneuploidies were diagnosed by conventional cytogenetic analysis. Of the 45 diagnoses of autosomal chromosome aneuploidies, pregnancy was terminated in 93.3%.

2479F An exploration of families’ experiences regarding a comorbid diagnosis of neurofibromatosis type 1 (NF1) and autism spectrum disorder (ASD) in their child: guiding screening and disclosure practices. L. Baret1,2, R. Hayeems3,1, C. Shuman1,2, M. Carter1,2, P. Kannel1,2, D. Chitayat3,2, A. Shugar2,1 1) The Hospital for Sick Children, Toronto, Ontario, Canada; 2) University of Toronto, Toronto, Ontario, Canada.

Introduction: Recent studies have identified an increased prevalence of autistic symptomatology in the NF1 population. Given these findings, it has been suggested that screening and diagnostic practices for young children with NF1 be expanded to encompass screening for ASD. Little is known about the attitudes of key stakeholders (parents of children with NF1 and ASD) regarding screening implementation.

Objective: We aim to explore the experiences of parents of children diagnosed with both NF1 and ASD and to probe their attitudes towards the potential implementation of early universal ASD screening for children with NF1.

Methods: We conducted a comprehensive retrospective chart review ascertained via the NF1 pediatric database at the Hospital for Sick Children. The chart review identified 22 children from 21 families with clinically confirmed diagnoses of both NF1 and ASD. With informed consent, we conducted 16 open-ended, semi-structured interviews with parents (73% response rate). Transcripts were analyzed for thematic patterns using qualitative content analysis.

Results: Most parents expressed relief after receiving an ASD diagnosis. Different meanings were conferred to the two diagnoses. The ASD diagnosis was perceived to be more closely tied the child’s identity, thereby conferring a greater impact on the parental psychosocial construct. The ASD diagnosis had both intrinsic and instrumental value for families as it also allowed them to access appropriate resources. In comparison, the meaning attributed to the NF1 diagnosis was less consequential. Interestingly, perceptions shifted if severe NF1 complications were experienced. Parents acknowledge the challenges associated with potential universal ASD screening; they fear a double label could provoke anxiety and present an additional barrier to adapting to NF1. Yet they almost unanimously support early ASD screening.

Conclusion: Implementing ASD screening for children with NF1 is likely to be beneficial for families if screening guidelines also address the psychosocial and medical impacts of an early comorbid diagnosis.

[Introduction] Thorough genetic counseling is essential for preimplantation genetic diagnosis and prenatal diagnosis. However, there are more than a few cases in which the time available for the diagnosis is limited. Here, we discuss and report the importance of timely counseling based on our experiences in two prenatally diagnosed cases. [Case 1] A 38-year-old unipara. An urgent hysterectomy was performed due to polyhydramnios and non-reassuring fetal status at the previous delivery. The newborn child died at one day old due to respiratory failure and was diagnosed clinically having congenital myotonic dystrophy. This result made us to suspect the same disease in the mother and the mother was followed up at our department of internal medicine. Two years later, she visited our hospital with a history of secondary amenorrhea and was diagnosed as being pregnant. On learning about her unexpected pregnancy, she was confused, as she had been told that a natural pregnancy would be difficult and she did not wish to have the baby. She was already 15-weeks pregnant. Because she was suspected to be a carrier based on the biochemical findings, she visited our hospital to seek counseling for her next pregnancy. While she and her husband were receiving counseling, her pregnancy was detected. [Results] Because in both cases, both the husband and wife sought a prenatal diagnosis, we set up an appointment for viii collection urgently at another institution and the gene analysis was performed. Prenatal diagnosis could be obtained in both cases. [Conclusion] In reproductive genetic counseling, some explanations about the next pregnancy is provided, however, the understanding usually becomes vague with time. In addition, we should keep in mind that a pregnancy is not always planned. It goes without saying that the independence of the client must be respected in genetic counseling, however, it is also important to establish a continuous communication and support system in reproductive genetic counseling in cases with time limitation.
Comparison of uptake of testing and psychosocial impact in pregnant and non-pregnant women offered carrier screening for fragile X syndrome.

Comparing uptake of testing and psychosocial impact in pregnant and non-pregnant women offered carrier screening for fragile X syndrome. S.Metcalfe1,2*,M.Martyn1,J.Emery1,J.Halliday1,2, S.Donath1,2, J.L.Vaugh1,2, J.Halliday1,2, S.Donath1,2, J.Cohen1,2, the FaXeS study team.

1) Murdoch Children’s Research Institute, Royal Children’s Hospital, Melbourne, Australia; 2) Dept Paediatrics, The University of Melbourne, Melbourne, Australia; 3) Dept General Practice, The University of Western Australia, Perth, Australia; 4) Fragile X Alliance and Centre for Developmental Disability Health Victoria, Monash University, Melbourne, Australia.

Offering population carrier screening to women for FXS identifies those at increased risk of having an affected child, and provides women with information about their own health risk. Concerns have been expressed around educational and counselling difficulties associated with the complex genetic and health information inherent in FXS carrier screening. Hence we examined decision-making, knowledge and attitudes in non-pregnant and pregnant women in the general population offered carrier screening for FXS.

Women were approached through general practice, obstetric or ultrasound clinics, and received written information and telephone pre-test counselling with consent. At home, women decided about testing and completed a questionnaire (Q1), which they returned in the mail with their buccal sample as appropriate. A premutation (PM) or grey zone (GZ) result was discussed by telephone and women offered genetic counselling; test-negative results were mailed. All women received a second questionnaire (Q2). A sub-group of women were interviewed either at the time of making the decision or after completing Q2; program evaluation interviews were conducted with clinic staff.

1237 women initially consented: 702 non-pregnant and 535 pregnant women. Excluding 80 women who subsequently withdrew from the study, 85% and 81% returned Q1 while 78% and 76% returned Q2 respectively. 71% non-pregnant and 59% pregnant women chose testing; 0.4% (n=3) received a PM and 2.0% (n=15) a GZ result. Overall, 85% had good knowledge (≥7/10 correct) at Q1. 77% non-pregnant and 68% pregnant women had positive attitudes towards the test. Pregnant women were less depressed and less stressed than non-pregnant women on the Depression Anxiety Stress Scale, with no differences between tested and non-tested.

Scores on this scale were in the normal range for all but a few women. The majority of women offered FXS carrier screening had good understanding despite the complex nature of the information, with minimal psychosocial impact. These may be related to the extensive prior field-testing of the written information and the pre-test counselling embedded within this study, important elements for consideration in screening programs. Overall, women supported the availability of being offered screening, although testing before pregnancy was preferred.

A study in contrasts: The effect of personal genomic testing on perceived risk of melanoma and lung cancer in the PGen Study.

D.A. Carera1, P. Kraft1, C.A. Chen1, L.A. Cupples1, T. Moreno2, J. Mountain2, J.S. Roberts3, R.C. Green4, the PGen Study Group.

1) Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115; 2) Data Coordinating Center, Boston University School of Public Health, 801 Massachusetts Avenue, Crosson 3rd floor, Boston, MA 02118; 3) Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston MA 02118; 4) Pathway Genomics Corporation, 4045 Sorrento Valley Blvd., San Diego CA, 92121; 5) 23andMe, Inc. 1390 Shorebird Way, Mountain View, CA 94043; 6) Department of Health Behavior and Health Education, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109; 7) Division of Genetics, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, 41 Avenue Louis Pasteur, Suite 301, Boston, MA 02115.

Background: In the Impact of Personal Genomics (PGen) study, we collaborated with Pathway Genomics and 23andMe to follow new customers for 6 months, linking individual genetic results with longitudinal survey data. Here we evaluate the impact of genetic results on perceived risk (PR) of two cancers with primarily non-genetic risk factors: lung cancer (PRL) and melanoma (PRM). We predicted that an elevated risk genetic result would be associated with an increase in PR and that the effect would be modified by baseline PR. Methods: PR was measured on a 5-point scale from much lower than average (1) to much higher than average (5) at baseline (pre-results) and 6 months post-results. Genetic results were based on common variants from genome-wide association studies and dichotomized as elevated risk or average/low risk. Linear regression for change in PR (6 months minus baseline) was performed, adjusting for baseline PR, age, sex, education, income, race, and company. Analyses stratified by baseline PR, and interaction terms between baseline PR and genetic result, were evaluated.

Results: Among 890 participants for whom data on lung cancer risk was available, the proportion rating themselves at above average risk was constant (12.9% at baseline; 11.4% at 6 months). Genetic result was significantly associated with change in PR_L (p<.0001). In those who received an elevated risk result, PR_L increased by an average 0.40 points on the 5-point scale; PR_L was unchanged in those with an average/low risk result. The effect of genetic result was greater in those who reported an above average risk at baseline than those who did not, but the difference was not significant. Data on melanoma risk was available for 509 participants, and the proportion rating themselves at above average risk for melanoma fell significantly from 48.5% at baseline to 35.2% at 6 months (p<.0001). Genetic result was significantly associated with change in PR_M (p<.0001), with no effect modification by baseline PR. PR_M was unchanged in those with an elevated risk result, but decreased by an average 0.61 points in those with an average/low risk result. Conclusions: The effect of genetic result on PR_L and PR_M is significant but modest. For lung cancer, an elevated risk result has a greater influence on PR than an average/low risk result, while the opposite is true for melanoma. This finding could be explained by consumers’ non-genetic risk factor exposure status, and further analysis is warranted.
2486F  Genetic testing for cardiovascular risk can promote better control over controlled risk factors during one year follow up in women. O.A. Makeeva, V.V. Markova. Inst of Medical Genetics, Tomsk, Russian Federation.

Aim of the study was to test a hypothesis that genetic testing for disease predisposition can have bigger impact on controlling existing risk factors for future disease than regular medical recommendation on improving life style. Methods and samples: To test the hypothesis two samples of relatively healthy overweight young adults were recruited during routine annual medical examination. First sample (N=95), 70% males, mean age 31±6 years, mean body mass index 29.3±3.5 underwent similar recommendation on body mass lowering and were offered to undergo genetic testing in respect to cardiovascular risk assessment. After informed consent obtained, blood samples were withdrawn and genotyping for several well known risk alleles completed. In a month period subjects were invited for genetic consultation and their risk for CVD communicated. In one-year period subjects from both samples were contacted to invite for the follow up visit. The information about their weight and BMI was collected. Results: Weight difference (difference between weight measures at the first and follow-up visits) was analyzed for two samples. Both samples reduced their weight to some extent. Mean weight difference for the group without genotyping was (-1.51±0.35) kg and for the group which undergone risk genetic assessment was (-2.29±0.45) kg which was not significantly different. When men and women were analyzed separately, a significant difference was observed, that in women the weight loss was higher in the subgroup who underwent genetic testing (-4.22±0.80 vs -0.82±0.55) kg, p=0.0006, while in men it did not differ (-1.50±0.53 vs -1.81±0.43) kg. In subjects who were genetically tested this association did not depend on a risk estimate (high, low, average risk of CVD). Risk estimates distribution factor versus CVD was 2.05 for women vs 1.33 for men, p<0.001. Together these findings suggest that women who tested positive for a BRCA1/2 mutation between 2 and 60 months prior to data collection were more likely to use genetic testing for the future cardiovascular risk assessment. Conclusion: Women tend to pay more attention for doctor's recommendation in respect to body mass control when this was followed by genetic risk assessment, but risk estimate itself was not the factor, which had influence. Men generally responded to doctor's recommendation for BMI reduction and genetic testing for the future health problems had no impact.

2487F  Disease Status and Genetic Testing Among Consumers of Two Personal Genomics Companies: Findings from the PGen Study. S.F. Meisels¹, J. Wardle¹, J. Mountain¹, T. Moreno¹, S.S. Kalia², J.S. Roberts³, R.C. Green⁴ for the PGen Study group. 1) Health Behaviour Research Centre, University College London, London, United Kingdom; 2)23andMe Inc; 3) Pathway Genomics; 4) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School; 5) Department of Health Behaviour & Health Education, University of Michigan School of Public Health.

Background: There is a widespread assumption that risk prediction is a major driver of consumer interest in genetic testing. However, an alternative motivation might be to find out whether existing diseases have a genetic etiology. While not one of our core hypotheses, our data analysis examined interest in genetic testing in relation to disease status using data from the PGen (Impact of Personal Genomics) Study, a prospective study in collaboration with two genetic testing companies, 23andMe and Pathway Genomics. Baseline data on current illness (e.g., somatognosics, and family history (FH)) was used to examine whether having a particular diagnosis predicted interest in receiving a genetic test result for that condition. Analyses were carried out for 12 conditions included in the genetic profile (Arthritis, Asthma, Heart Disease (HD), High Cholesterol, Diabetes, Obesity, Breast Cancer, Skin Cancer, Multiple Sclerosis, Ulcerative Colitis (UC), Macular Degeneration and Bipolar Disorder). For each condition, we tested whether having a diagnosis was associated with being more likely to express ‘high’ interest in receiving genetic test feedback for the same condition. Results: 1801 participants provided data on the baseline survey. Their mean age at baseline was 46 years (SD 15.6 years), they were predominantly white (89.4%), just over half were women (60.5%), and nearly half were at least college educated (47.1%). About 80% of them had a diagnosis of at least one of the 12 conditions studied. When comparing test results to those without a diagnosis, test results ranged from 28% (UC) to 68% (HD). While each condition, logistic regression analysis revealed that after adjusting for demographic and personal variables and the company offering the test, having a diagnosis of the condition increased the odds of genetic testing. This was true for the 12 conditions examined. These findings suggest that some consumers might seek genetic testing not only for its predictive value, but also to explore the etiology of a condition they already had. This motivation has so far not been described in the debate about clinical utility of consumer based genetic testing, but should be considered as an important factor influencing decisions to seek testing.

2488F  Pathways to and through genetic testing and cancer aged 18-25. L.M. Hoskins¹, A. Werner-Lin². 1) Clinical Genetics Branch, National Cancer Institute, National Institutes of Health; 2) School of Social Work, New York University.

Background: Much of the extent the literature on the psychosocial aspects of BRCA1/2 mutation testing aggregates mutational carriers of all ages in recruitment, analysis, and interpretation. This strategy does not adequately address the needs of the youngest genetic testing consumers, i.e., women aged 18-25. Despite low absolute cancer risk estimates before age 30, BRCA1/2 mutation-positive women aged 18-25 feel vulnerable to a cancer diagnosis and feel compromises in a management quandary because the clinical utility of screening and prevention options are not yet well defined for such young carriers. This presentation aims to demonstrate the unique developmental, relational and policy influences, as well as the challenges, experienced by very young BRCA1/2 mutation-positive women as they complete genetic testing and initiate cancer risk management. Methods: This study integrated qualitative data from three independent investigations of BRCA1/2-positive women recruited through cancer risk clinics, hospital-based research centers, and online organizations. We present three representative cases, selected from this larger study of 32 BRCA1/2 mutation-positive women aged 21-25 who tested positive for a BRCA1/2 gene mutation between 2 and 60 months prior to data collection. Findings: The first case describes the maturation of a young woman whose family participated in a national cancer registry, the second addresses a young woman who completed genetic testing after learning that her unaffected father was a mutation carrier. The third case highlights a young woman parentally bereaved in childhood who presented for genetic testing due to intense family history concern. Together these cases demonstrate that young women aged 18-25 are challenged to reconcile their burgeoning independence from their families with risk-related support needs. Conclusions: Young women who carry deleterious mutations in BRCA1 and BRCA2 face not only increased short-term and lifetime risk of developing cancer, but also a unique set of challenges related to managing cancer risk during an already demanding phase of life. Loved ones acting in ways meant to care for these young women may inadvertently apply pressure, compelling family support dynamics, and self-nurturing decision-making. Ongoing support from competent healthcare professionals will enable these young women to remain informed and receive objective counsel about their risk-management decisions.

2489F  Moving Up Mastectomies: Emerging Adulthood, Surveillance Fatigue and the Affordable Care Act of 2010. A. Werner-Lin¹, L. Hoskins². 1) Silver School of Social Work, New York University, New York, NY; 2) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.

Significance: Increasingly, women aged 18-25 pursue genetic testing to identify a BRCA gene mutation, despite low absolute cancer risk and the lack of evidence-based options for prevention and risk reduction. Complicating risk management decisions are a family context that clouds autonomous election of risk management strategies with prolonged grief and uncertainty. Further, the Affordable Care Act created a public policy context that potentially reduces barriers to genetic testing, enabling current and future BRCA1/2 mutation-positive women aged 18-25 to have genetic services covered. Methods: This study integrated qualitative data from three independent investigations of BRCA1/2-positive women recruited through cancer risk clinics, hospital research centers, and online organizations. Data were originally subsampled as a part of each study cohort. 26 participants aged 18-25 at data collection were re-analyzed. In 2011, 6 BRCA1/2-mutation positive women aged 21-25 were recruited at the meeting of a web-based advocacy group. All 32 women tested positive for a BRCA1/2 gene mutation between 2 and 60 months prior to data collection. Investigators used grounded theory and interpretive description to conduct both within and cross-study analysis. Findings: Women from all three cohorts expressed the need for clarity in recommendations for screening and prevention before age 25, especially with consideration of early and young-age independent decisions. Complicating risk management decisions are a family context that clouds autonomous election of risk management strategies with prolonged grief and uncertainty. Further, the Affordable Care Act created a public policy context that potentially reduces barriers to genetic testing, enabling current and future BRCA1/2 mutation-positive women aged 18-25 to have genetic services covered. Methods: This study integrated qualitative data from three independent investigations of BRCA1/2-positive women recruited through cancer risk clinics, hospital research centers, and online organizations. Data were originally subsampled as a part of each study cohort. 26 participants aged 18-25 at data collection were re-analyzed. In 2011, 6 BRCA1/2-mutation positive women aged 21-25 were recruited at the meeting of a web-based advocacy group. All 32 women tested positive for a BRCA1/2 gene mutation between 2 and 60 months prior to data collection. Investigators used grounded theory and interpretive description to conduct both within and cross-study analysis. Findings: Women from all three cohorts expressed the need for clarity in recommendations for screening and prevention before age 25, especially with consideration of early and young-age independent decisions. Complicating risk management decisions are a family context that clouds autonomous election of risk management strategies with prolonged grief and uncertainty. Further, the Affordable Care Act created a public policy context that potentially reduces barriers to genetic testing, enabling current and future BRCA1/2 mutation-positive women aged 18-25 to have genetic services covered.
2492F Is it ‘just like any other test’? Parents experiences with array cGH in pediatrics. R. Babul-Hirji, N. Hoang, R. Hayeem, R. Weksberg, C. Shuman. Division of Clinical & Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada.

Background: Understanding families’ perceptions of the value and complexity of array cGH testing will enhance its delivery in the clinical setting. This information will also serve to inform the implementation of next-generation sequencing in pediatrics. Methods: Sixty-nine families who pursued array cGH testing for children with congenital anomalies +/- delay were invited to participate in semi-structured interviews. Fourteen families accepted, 19 declined, and 44 have not responded. Of the 13 interviews conducted to date, parents represented diverse ethnic backgrounds and the majority were well educated. The age of the child at time of array cGH ranged from 3 days to 21 years with an average age of 6.8 years. Six had an uncertain result, 4 had a pathogenic result, and 3 had a benign result. Qualitative analytic techniques were consistent with interpretive description. Results: Preliminary analysis indicates the following themes: (i) microarray test is not readily distinguished from other genetic tests; (ii) uncertainty is experienced regardless of result type and responses to this uncertainty include both hope and frustration; (iii) families articulate a personal utility of array results even if they do not perceive specific utility for medical management, access to support services, and family planning. Conclusion: Our finding that families do not distinguish among types of genetic tests should be considered in the context of pre-test genetic counseling. While emerging frameworks for genome-based testing emphasize extensive counseling about the nature and possible outcomes of testing, parents appear to focus on the rationale for testing and the potential for additional diagnostic information. Finally, emerging genetic counseling approaches should be informed by our findings that receiving uncertain information does not appear to compound the intrinsic uncertainties associated with caring for a child with a rare or undiagnosed genetic condition, but this merits further research.

2491F A Study of Perception of Health Problems in Patients with Prader-Willi Syndrome by Their Caregivers and the Caregivers’ Health Care Behaviors. W. Khunin, P. Tampowong, D. Wattanasinchaiagoon. Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

Background: Prader-Willi syndrome (PWS) is a rare genetic disorder characterized by infantile hypotonia, lethargy, feeding difficulty, poor weight gain, followed by excessive eating, morbid obesity, obstructive sleep apnea, endocrinopathies, cognitive impairment, and multiple behavioral problems later in life. Treatments include multidisciplinary medical approach as well as behavioral and environmental modification by caregivers at home, which require excessive understanding and patience. The present study is aimed to explore perception of health problems in patients with PWS by their caregivers and the caregivers’ health care behaviors. Methods: This is a descriptive study. Eligible subjects were primary caregivers of patients with confirmed PWS. The caregivers answered questionnaire by mail, or by phone or face-to-face interview. The questionnaire questions comprise demographic data of the caregivers and the children, health problems of the patients, and the caregiver’s health care behavior which include regularity of medical care and treatment. Results: The major reason of loss follow-up; promotion of the children to take part in social activities; registration for disability benefit card; urgent needs for help; impact to the family members by having a child with PWS. We also asked the effect of having patients with PWS onto their family planning, acceptance for prenatal diagnosis, and choosing option for termination of pregnancy if the fetus was found affected prenatally. Results and Conclusion: Thirty-one caregivers participated the study, including 29 mothers and 2 fathers. The age ranges from 2 to 25 years (mean age 9 years). In regards to health problems, the frequency of health problems for caregivers perceived are as followed: obesity 24/31, sleep problem 13/14, short stature 11/14, undescended testes 8/9, over-compulsive 13/29, developmental delay 21/31, regular follow-up 25/31, register for disability benefit card 19/31, having patient with PWS affect their physical health 15/31 and psychosocial/emotional health 29/31. In term of acceptance for prenatal diagnosis (PND) for the next pregnancy, 28/31 accepted PND and 15/28 would op termination of pregnancy. In conclusion, the study add the value of caregivers’ perception which will be essential for tailoring patient care and for helping the caregiver to cope with stress and get support for themselves as well.
2494F  
Cancer Genetics Referral Patterns of Physicians and Patient Socio-Demographics. J. Cohn, W. Blazey, S. Koehler, B. Laurent, V. Chan, M. Jung, D. Tan, B. Krishnaachari. NYIT College of Osteopathic Medicine, Old Westbury, NY.

BACKGROUND: Genetic testing can aid with the selection of cancer screening and preventive health services in high-risk patients. Studies demonstrate that physicians are more likely to order genetic testing based on patient inquiry rather than on recommended guidelines. Literature also shows that there are racial disparities in patient use of genetic services that do not appear to be explained by differences in socioeconomic factors. The association between patient use of genetic services and race is not well studied.

METHODS: The goal of this study was to investigate racial and socio-economic disparities in cancer genetic risk assessment and referral patterns in clinical practice. A survey was administered to physicians (primarily primary care physicians) evaluating physicians’ awareness of the racial and socio-economic details of their patients, their genetic risk assessment and referral patterns.

RESULTS: 139 physicians completed the survey. 26.6% (n=37) of the physicians reported conducting frequent genetic risk assessment and 44.6% (n=62) reported frequently referring patients for genetic services. 13.8% (n=19) reported that at least 25% of their patients were African American, and 26.8% (n=37) did not know what percentage of their patients were African American. Additionally, 35.8% (n=59) did not know what percentage of their patients were Latino/Hispanic and 34.9% (n=58) did not know what percentage of their patients were classified as living at poverty levels. Those who conducted frequent risk assessment were more likely to state that they would refer a non-English speaker for genetic testing regardless of language barriers (OR=8.34, 95% CI: 1.08-64.69, p<0.05). Those who frequently referred patients for genetic services were more likely to state they would refer a Medicaid patient for genetic testing even when they thought the testing costs would not be covered (OR=2.44, 95% CI: 1.20, 4.99, p<0.05). CONCLUSIONS: Most research has focused on disparities in the perspective of patient preferences and behaviors. In this study, about a third of physicians surveyed reported not knowing the socio-demographic breakdown of their patient populations and may not be conscious that some of their patients come from socio-demographic groups less likely to pursue genetic services on their own. Further research must focus on clinical practice and physician behaviors.

2495F  

1) Laboratory Corporation of America, Research Triangle Park, NC, United States; 2) LabCorp/Dynacare, Seattle, WA, United States; 3) LabCorp/DynaGene, Houston, TX, United States.

Prenatal microarray analysis has been employed for a number of years, but its population use has exponentially increased over the past two years, especially with the publication of the NICHD study. For 1 1/2 years we have utilized an Affymetrix SNP microarray platform to study over 8,000 prenatal patients. This work has highlighted some of the counseling issues that have become important to us that are the result of the following: The potential for genetic risks associated with structural chromosome abnormalities subsequently studied by the array had parental studies in 45% of the cases. This work has highlighted a number of counseling issues including: (1) Need for pre-test counseling delineating potential risk of the chromosomal abnormalities identified; (2) A majority of the patients who underwent microarray testing in this study were referred by obstetricians who identified abnormalities detected prenatally; (3) For patients with abnormal deletions detected by microarray, we were unable to determine prenatal risks for genetic counseling purposes; (4) Chromosome analysis for microdeletion syndromes associated with autism and other abnormalities. This work has highlighted the potential utility of microarray technology for genetic counseling programs to help prepare new graduates for the complexities involved; and (7) Critical need for trained laboratory genetic counselors to assist in the dissemination of the results to providers.

2496F  
Offspring Risk Perceptions: Adolescents and Young Adults with Congenital Heart Disease Agree with their Parents (and both are wrong). S.M. Fitzgerald-Butt, A. K. FM. Fry, A. N. Zaidi, C.A. Gerhardt, V. Garg.

1) Center for Cardiovascular and Pulmonary Research, The Research Institute at Nationwide Children’s Hospital, Columbus, OH; 2) The Heart Center, Nationwide Children’s Hospital, Columbus, OH; 3) Center for BioBehavioral Health, The Research Institute at Nationwide Children’s Hospital, Columbus, OH; 4) Department of Pediatrics, College of Medicine, The Ohio State University, Columbus, OH.

As individuals with congenital heart disease (CHD) now typically survive into adulthood, there is an increasing demand for reproductive counseling. This is pertinent concern. Parents are the typical source of medical information through childhood but adolescents and young adults (AYA) must have sufficient knowledge to assume responsibility for their medical care. Therefore, we examined offspring risk perceptions alongside parental perceptions in adolescents and young adults with CHD to determine what level of agreement exists with their parents.

RESULTS: Participants included 249 AYA, 15-25 years old with structural CHD (mean age=19.0 years, 54% female, 85% Caucasian) and 179 parents (mean age=47 years, 34% male 94% Caucasian) who were recruited from an outpatient cardiology clinic (85% AYA consent rate). All participants were asked to rate the likelihood of their child’s CHD and completed measures of demographics, genetic knowledge, perceived CHD severity, and perceived risk for an offspring with CHD. CHD complexity was rated as simple, moderate, or great. Categorical perceptions of risk were similar in AYA and parents with only approximately a third of participants rating CHD as simple. There were two-three-fold increases in endorsed average or lower risk. The perception of being in the high risk category was associated with higher genetic knowledge (p<0.001) and higher perceived CHD severity in a hypothetical baby (p=0.001) in both AYA and parents. In a multiple regression model, having a household income (p=0.016) was associated in just the AYA and perceived future heart defect severity (p=0.016) was associated in just the parents. The ratings of risk on a continuous scale were extremely variable and remarkably high among the AYA and parents, ranging from 0-100% with a mean of 36.8% (SD=24.3%) and 34.1% (SD=23.5%), respectively, with a mode of 50% and increased selection of 25% and 75% in both groups. The majority of AYA and parents have an inaccurate categorical risk perception and their continuous risk perception is significantly different and either lack of or inaccurate knowledge of risk. These results highlight the need to provide accurate offspring risk information to both AYA and parents, possibly while also providing additional genetic and numeracy education.
2498F  
A Two-Stage Approach to Genetic Risk Prediction of Breast and Ovarian Cancer.  
P.M. Atienza1,2, J. Chipman3, K. Hughes4, C.I. Ames4, B. Arun2, C. Richardson1, S. Biassou3.  
1 School of Public Health-Biostatistics, University of North Texas Health Science Center, Fort Worth, TX; 2) Alcon Research, Ltd., Fort Worth, TX; 3) Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, MA; 4) Massachusetts General Hospital, Boston, MA; 5) Department of Community and Family Medicine, Geisel College of Medicine, Dartmouth College, Hanover, NH; 6) Department of Breast Medical Oncology and Clinical Cancer Genetics, University of Texas M.D. Anderson Cancer Center, Houston, TX; 7) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 8) Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX.

BRCAPro is widely used in genetic counseling to identify patients at high risk for hereditary breast or ovarian cancer. However, it requires extensive information on the patient's family history, which is a hindrance for its use in many health care settings. Recently, we proposed simplified versions of BRCAPro for settings that do not require exhaustive genetic counseling (Breast Cancer Research and Treatment, 139:571-579, 2013). As simpler versions lead to loss in accuracy as compared to BRCAPro, here we propose a two-stage approach to balance the tradeoff between the amount of information used and the accuracy achieved in assessing the risks. In the first stage, only a limited amount of family history will be analyzed using simpler versions of BRCAPro. If the risk at this stage is sufficiently high, BRCAPro, with exhaustive family information, will be used in the second stage to potentially achieve more accurate predictions. We consider four first stage tools: (1) BRCAPro-Plus uses information on relatives up to the second degree for those affected with breast cancer and/or ovarian cancer only; (2) BRCAPro-Plus additionally imputes the ages of unaffected relatives based on an external data source; (3) BRCAPro-Simple imputes the family structure and the ages of unaffected relatives up to the second degree; (4) BRCAPro-Degree uses information on only the first-degree affected relatives. The second stage is BRCAPro for all these tools.

We develop a methodology to evaluate the overall performance of a two-stage approach. In particular, we compute sensitivity, specificity, and predictive values and area under the ROC curve (AUC) of a two-stage approach by considering different combinations of the first and second stage cutoffs. We also evaluate the clinical implications of using a two-stage approach. We use 2,713 probands from seven sites of the Cancer Genetics Network and MD Anderson Cancer Center to evaluate our approach. The proposed two-stage approach has comparable discrimination and calibration as BRCAPro if the latter is applied to all probands. The overall AUCs are 0.782 (BRCAPro-Plus), 0.783 (BRCAPro-Simple), and 0.775 (BRCAPro-Degree) while the AUC for BRCAPro is 0.783. Also, the two-stage approach substantially reduces the genetic counseling and testing burden. Thus, we conclude that this approach can be adapted for genetic risk prediction of breast or ovarian cancer on a large scale.

2499F  
Clinical implications of Variants of Unknown Significance in Chromosomal Microarrays in pregnancies in Israel.  
H. Yonath1,2, S. Rienstein1, J. Shamay1,2, M. Berkenslady1, M. DiCicco1, S. Eisenberg-Brazili1, N. Goldstein3, M. Frydman1,2, E. Pras1,2.  
1 Inst Human Gen, Sheba Med Ctr, Ramat Gan, Israel; 2) Sackler Medical School, Tel-Aviv University, Tel-Aviv, Israel.

Background: One of the main concerns in finding a Variant of Unknown Significance (VUS) in Chromosomal Microarray (CMA) during pregnancy is the implication it would have on the decision to continue or terminate it. Many couples in Israel have a low tolerance for abnormalities detected during pregnancy and prefer to terminate a pregnancy if an abnormality is found. The goal of this study was to evaluate the decision made by the parents regarding pregnancies in which a VUS was found, in a large genetic institute in Israel. Methods: 305 prenatal CMA cases were evaluated. The arrays were oligonucleotide specially targeted prenatal arrays (Agilent arrays designed by Medical Genetics Laboratory at Baylor College of Medicine). The significance of the variants was determined based on local guidelines that are very similar to the ones described by Wapner et al (1). Parental CMA were performed in order to determine whether the finding is inherited or de novo. Genetic counseling was provided to the couples. Results: VUS were found in 36 cases (12%). About half of them were inherited. And the indications were mainly abnormal US and parental concern. None of these pregnancies were terminated. Conclusions: Though there is a high concern that VUS in CMA would lead to the termination of pregnancy, our results show that if there is no clinical data that suggests that the abnormality is pathogenic, and with the proper genetic counseling, the results do not lead to termination of the pregnancy. References: (1) Chromosomal microarray versus karyotyping for prenatal diagnosis.Wapner RJ et al. N Engl J Med. 2012.

2500F  
A gift to the children - genetic testing at the end of life.  
M. Mikhael1, K. Stears2.  
1 InVitae Corporation, San Francisco, CA; 2) St. Vincent's Hospital, Billings, MT.

INTRODUCTION: Recent publications discuss targeted testing for patients dying of ovarian cancer. Less is known about end of life testing for a patient with a complex history that does not directly point to a specific genetic disorder. Here we present the case of a patient who was seen for genetic counseling in order to learn what genetic mutations she may pass on to her descendants. The patient was consented for and tested for 262 genetic conditions by InVitae Corporation. CASE REPORT: The patient is a 81-year-old terminally ill woman with a history of peripheral vascular disease, pulmonary emboli, stroke, melanoma, breast cancer, calcium deposits, spinal stenosis, hyperlipidemia, hypertension and peripheral artery disease. She has three children. She has a family history of vascular disease, liver disease and cancer in first degree relatives and sudden cardiac death in a second degree relative. She did not have any previous genetic testing. The patient was interested in testing to provide carrier information for her descendants. Using the InVitae Assay, the patient tested for 262 conditions, including cardiac arrhythmias, a 50 gene cancer panel, clotting disorders and multiple other inherited conditions. A homozygous mutation for HFE related hereditary Hemochromatosis was found. The results prompted an additional Hemochromatosis evaluation, which was negative. DISCUSSION: The testing provided relevant information to the patient's children about the risk of Hemochromatosis and the reduced relative risk of inherited arrhythmias, clotting disorders and cancer predisposition. Broad genetic testing can be used as a supplement to DNA banking in a patient with a complex personal and family history. Issues around obtaining informed consent and patient confidentiality are discussed. A description of the genetic counselor's personal experience with this case is also detailed.
2501W

Using the Cleveland Clinic Score to predict for germline PTEN mutations in the analytical algorithm of Cowden syndrome: a cost-effectiveness analysis. N. Gross stheit 1, N. Sood 2, 3, M. S. Messer. 1) Global Biotech Consulting Group (GBCG), Mexico; 2) Genómica y Bioeconomía AC, Mexico; 3) Mexican Health Foundation, Mexico; 4) Harvard School of Public Health, Boston, MA.

Background: Cowden syndrome (CS) patients with germline PTEN mutations are at increased risk of benign and malignant features. Our group has developed an online risk calculator, the Cleveland Clinic Score (CC Score) to determine the likelihood of a cowden syndrome patient having a PTEN mutation.

Methods: The higher the CC score, the higher the probability of an underlying germline PTEN mutation. CC Score of 10 corresponds to an estimated 3% pathogenic PTENgermline mutation risk. Issues that affect clinical testing include sensitivity and specificity of the test, the benefit to the patients and the possible negative ramifications of the results, and the cost. The aim of this study was to assess the cost-effectiveness of using different CC Score thresholds to improve the yield of the diagnostic algorithm of CS. Methods: Data from an existing multicenter prospective study in which 3541 probands satisfying relaxed CS clinical criteria were used. PTEN mutation scanning, including promoter and large deletion analysis, was performed for all subjects.

A decision model was developed to estimate the number of mutation carriers and the incremental costs of alternative case-finding methods for detecting PTEN mutation carriers. Strategy 1 involved PTEN testing of all individuals meeting relaxed CS clinical criteria. Strategy 2 involved PTEN mutation testing only in individuals with CC Scores ≥ 10. Incremental costs of using different CC Score thresholds (15, ≥20) for detecting PTEN mutation carriers were estimated. For each strategy, the number of PTEN mutations detected, mutations detected, missed, false positive results were computed. One-way sensitivity analysis was performed to assess robustness of estimations. Results: Pathogenic PTEN mutations were identified in 250 individuals (7.1%). Sensitivity for detecting PTEN mutation carriers using CC Score cutoff of 10 (strategy 2) was 83% and specificity was 71%, missing 49 patients with germline PTEN mutations. Average cost per PTEN mutation detected using Strategy 1 was $21246 compared with $8455 for Strategy 2. The incremental cost per additional case detected using Strategy 2 (all patients tested) was $73714. Conclusions: Use of clinical risk calculators such as the CC Score is a cost-effective pre-screening method in the selection of patients for PTEN germline analysis when CS is suspected.

2502W

Genetic tests evolution in the Genomic era. Is cost evaluation a relevant factor in health care planning? D. Coviello, C. Lanza, A. Seri, M. Parodi, P. Casale, A. Fabbrini, S. Casati, M. Esposto. 1) Laboratory of Human Genetics, E.O. Ospedali Galliera, Genoa, Italy; 2) Management Control Unit, E.O. Ospedali Galliera, Genoa, Italy; 3) UNIAMO, Italian Patient Federation on Rare Diseases, Venice, Italy.

In the last thirty years, medical genetics has improved molecular diagnosis according to the evolution of gene technology, but it is really in the last five years that the new revolution of next generation sequencing (NGS) has open a new window not only on research but also on diagnostic testing. Laboratories performing routine genetic testing started their activity using mainly high throughput technology. The use of NGS platform and the laboratory activities are completely changed: the number of known disease genes has increased, automation is applied also to the laboratory, cerification and accreditation procedures of genetic services are in place, information and awareness among public and all stakeholders are crucial. Purpose of the study: both public National Health systems (NHS) and private laboratories or companies need to deal with costs of genetic testing performed with traditional methodologies or with the new strategies derived by the introduction of NGS. Since in this field there is still not clearly defined how many genetic tests have enough clinical utility, the investigation of their costs could be a way to establish a correct public health policy. Methods used: Activity-based costing (ABC) is a methodology that identifyes activities and assigns the cost of each activity taking into account resources, all products and services according to the actual consumption by each. This methodology has been applied to the steps of diagnostic processes of our genetic laboratory. Summary of results: The analysis of the activities has identified a number of indicators to assess the workload for every activity deriving useful information on the costs of test providers, identifying the part of the analysis has evaluated the proportion of material used for each activity (test or portion of the process). All these parameters have been incorporated into a software able to split all the lab costs (personnel, material and non material) and to assign costs to every test provider, in comparing the performance of the same laboratory in subsequent years or to make a priority list of genetic/genomic tests to provide, taking into account costs/benefits data.

2503W

Innovation in genomic medicine to realize the bioeconomy in Mexico. S. March 1, 2, F. Valdez-Ortega 1, 2, G. Soberon 1, 2, J. Frenk 1, 2, G. Jimenez-Sanchez 1, 2, 3, 4, 1) Global Biotech Consulting Group (GBCG), Mexico; 2) Genómica y Bioeconomía AC, Mexico; 3) Mexican Health Foundation, Mexico; 4) Harvard School of Public Health, Boston, MA.

A bioeconomy is one based on the use of research and innovation in the biological sciences, rate economic activity and public benefit. It is a large and rapidly growing segment of the world economy that provides substantial public benefit. The growth of the bioeconomy is due in large part to the development of genomics. In Mexico, genomic sciences are vigorously developing, particularly in the use of genomic medicine. Genomic medicine and health are being developed in the developing world (D Kumar Ed, Oxford University Press, 2012). To stimulate strategic synergies for innovation in genomics, we established Genomica y Bioeconomia, a non-profit organization led by experienced leaders in genomics, health policy, economy, business development, legal and social sciences (www.genomicaybioeconomia.org). Our initial results include a comparative analysis of knowledge-based economic initiatives in emerging economies including China, India, and Brazil. This led to specific recommendations as to how an emerging economy like Mexico can orient major genomic programs into innovation and economic development (Jimenez-Sanchez G, et al. Genomics and Bioeconomy: A window of opportunity for economic growth in Mexico, COLMEX-COLNAL, 2012). Genomic medicine is one of the key areas where previous work and novel developments in science and industry make it necessary to establish the grounds for implementation of genomic applications into medical practice. This is the case for pharmacogenomics, where a genetic diagnosis is not enough to make a medical decision. Agencies that make it important to develop cost-benefit analyses for some of the most commonly used tests. We initiated an economic evaluation of the Warfarin sensitive test in the context of net present values for Mexico. Initial evidence indicates that this test could be used in the public health arena. Based on current technologies and the knowledge of genetic variation in the Mexican populations, we established a joint effort with the Mexican Government to develop a regulatory framework for pharmacogenomics and the infrastructure to support the COPEPRIS (FDA equivalent) laboratory facilities. Results from this initiative include a nation-wide network to establish synergies between academy, industry and government to develop innovation in genomics oriented to public health challenges of the Mexican population.

Supported in part by CONACYT.

2504W


Following Angelina’s Jolie statement on her BRCA1 status, awareness of possible BRCA mutation in families was considerably raised. Less than a year after that the statement, the CEO of deCode Genetics announced that from its research, the company had data on all BRCA2 mutation carriers in Iceland. This information could be a huge benefit for other cancers. Use of the database and registry, based on presumed consent, enables for very fast and accurate workup and risk assessment. The CGC unit at the National Landspitali University Hospital has been in operation since December 2006. One genetic counselor and one part-time oncologist work at the unit. In 2011 and 2012, 230 patients have come for counseling, the majority of BRCA2 families are thought to have been identified and all BRCA1 families. As families can be traced very far back and the initial BRCA2 carrier has been identified from the 16th century, the number of cases is always rising. When a BRCA mutation is found, the carrier gets an information letter about the mutation and the genetic counseling service to distribute to his family. We describe how a very small CGC unit coped with a very big workload and complex counseling issues in a very short time.

Background: The Hereditary Non-polyposis Colorectal Cancer (HNPPC) study is a clinical trial conducted in an integrated health-care system—Kaiser Permanente Northwest (KPNW). Our goal is to determine whether universal screening for Lynch Syndrome among all newly diagnosed colorectal cancer cases increases the use of genetic counseling, compared with the current practice of physician-based referrals or self-referrals to medical genetics. To this end, we interviewed health plan leaders and providers to better understand requirements for successful implementation of universal screening. Methods: We investigated factors that might hinder or facilitate implementation of universal screening for Lynch Syndrome in all newly diagnosed colorectal cancer patients. We conducted fifteen semi-structured interviews with KPNW leaders and key staff in pathology, oncology, medical genetics, gynecology, surgery, and laboratory services. Using NVivo software, we applied thematic analysis to the interview audio transcripts. Results: Three major themes emerged: 1) departments that should be represented in planning and implementation decisions; 2) key decisions to be made include: patient selection criteria, consent protocols, choice of Microsatellite Instability (MSI) or Immunohistochemistry (IHC) testing—or both—for initial testing, laboratory selection, and whether to include patients with endometrial cancer in the screened population. These decisions will influence several potential barriers and issues. Justification of screening decisions based on quantitative benefit analysis will be necessary, as will tracking mechanisms for orders, results, and surveillance follow-up. Training and workload burden were common barriers. "The idea of a genetic counselor—there's just none in some departments." Clinicians consistently supported systematic testing, suggesting a broader belief within the organization that it 's the right thing to do.' Conclusion: Successful implementation of effective systematic screening for Lynch Syndrome is contingent on cross-disciplinary collaboration and coordination. Program development should include assessment of alternate screening strategies. Barriers must be addressed to achieve a process that is both well integrated and cost effective.

Engaging health professionals in evaluations of emerging genomic technologies. C. G. C. G. C. (UT); 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Department of Pediatrics, University of Washington, Seattle, WA; 4) Department of Family Medicine, Mount Rainier Medical Center, Seattle, WA; 5) Department of Health Care Informatics, University of Washington, Seattle, WA; 6) Department of Public Health, University of Ottawa, ON, Canada.

Background: My46 (http://www.my46.org) is an innovative web-based tool designed to enable self-guided management of genetic testing results. This online tool empowers individuals and families to make decisions about what genetic test results might be of utility to them and their children, and why they might or might not be useful. A pilot study was conducted that aimed at understanding the preferences and needs of individuals about genetics, risk of transmission, and medical conditions thereby providing genetic counselors more time to focus counseling sessions on interpreting genetic results and exploring their impact. To facilitate this aim, My46 has developed an open-source "Learning Center" that includes standardized summaries of a wide range of medical conditions written in patient-friendly language ("Trait Profiles"); key considerations to think about before receiving genetic testing results ("What you should know about genetic testing"); an introduction to human genetics with an extensive glossary of terms; and links to external resources for families and healthcare professionals ("More Resources"). The Trait Profiles can be searched by trait name (e.g., sickle cell disease) and are organized into seven categories (disease, risk, carrier status, medication response, genetic syndromes, metabolic disorders, newborn screening conditions, and ancestry). The Trait Profiles include information on characteristics of the condition, diagnosis/ testing, management/ surveillance, genetic counseling, special consideration, resources with direct links to GeneReviews™, and key references. All Trait Profiles have been reviewed as "Patient-friendly" by a genetic counselor, written by a genetic counselor, edited by the My46 team for readability, and reviewed by a clinical geneticist. The goal is to write a Trait Profile for each condition for which genetic testing is available. To date, the My46 team has received over 1000 Trait Profiles, written by genetic counselors in a single repository, greatly benefiting vast numbers of patients. Collectively, the My46 Learning Center is a valuable resource for the genetic counseling community and will enhance the delivery of genetic counseling services.

2506W


Background: The Hereditary Non-polyposis Colorectal Cancer (HNPPC) study is a clinical trial conducted in an integrated health-care system—Kaiser Permanente Northwest (KPNW). Our goal is to determine whether universal screening for Lynch Syndrome among all newly diagnosed colorectal cancer cases increases the use of genetic counseling, compared with the current practice of physician-based referrals or self-referrals to medical genetics. To this end, we interviewed health plan leaders and providers to better understand requirements for successful implementation of universal screening. Methods: We investigated factors that might hinder or facilitate implementation of universal screening for Lynch Syndrome in all newly diagnosed colorectal cancer patients. We conducted fifteen semi-structured interviews with KPNW leaders and key staff in pathology, oncology, medical genetics, gynecology, surgery, and laboratory services. Using NVivo software, we applied thematic analysis to the interview audio transcripts. Results: Three major themes emerged: 1) departments that should be represented in planning and implementation decisions; 2) key decisions to be made include: patient selection criteria, consent protocols, choice of Microsatellite Instability (MSI) or Immunohistochemistry (IHC) testing—or both—for initial testing, laboratory selection, and whether to include patients with endometrial cancer in the screened population. These decisions will influence several potential barriers and issues. Justification of screening decisions based on quantitative benefit analysis will be necessary, as will tracking mechanisms for orders, results, and surveillance follow-up. Training and workload burden were common barriers. "The idea of a genetic counselor—there’s just none in some departments." Clinicians consistently supported systematic testing, suggesting a broader belief within the organization that it’s the right thing to do. Conclusion: Successful implementation of effective systematic screening for Lynch Syndrome is contingent on cross-disciplinary collaboration and coordination. Program development should include assessment of alternate screening strategies. Barriers must be addressed to achieve a process that is both well integrated and cost effective.

2507W


Background: MyFamilyHealthHistory (MyFHH) is a family history collection tool developed by Cleveland Clinic (CC) for use in oncology settings. MyFHH is a CC facility improvement initiative to increase the efficiency of cancer family history collection without introducing care disparity. We have previously shown patients who complete MyFHH are more likely to attend cancer genetics clinic appointments (93% vs. 63% attendance in those not completing). We thus investigate determinants of MyFHH uptake. Methods: Between Aug 2009 and Sept 2012, 1161 patients scheduled appointments triggering MyFHH invitation. Cancer-specific family history is obtained for first and second degree relatives. Univariable/multivariable analysis of the association between completion status and personal history of neoplasm, sex, age and socioeconomic status (SES) was performed. Demographic data was extracted from the electronic medical record; median household income by ZIP Census Tabulation Area was used as a proxy for SES. Results: 359 (31%) completed MyFHH. Of 1161 invited, 877 (76%) had a personal diagnosis of neoplasm, 1002 (84%) were female, and 594 (77%) were age >65 yrs. In univariable analysis, we did not find evidence of a difference in the odds of completing MyFHH based on personal diagnosis of neoplasm (p=0.353). Men showed a reduction in the odds of completing MyFHH as compared to women (OR 0.69; 95%CI 0.46, 1.04), although the difference between the groups was not statistically significant (p=0.097). Notably, individuals age >65yrs were significantly less likely to complete MyFHH as compared with those age <65yrs (OR 0.47; 95%CI 0.31, 0.71; p<0.001).

2505W


Background: Engaging health professionals in evaluations of emerging genomic technologies into mainstream health care requires their acceptance and adoption by professionals across many disciplines and specialties. It is well established that interventions of known effectiveness are not used as intended, and that ineffective or harmful interventions continue to be used inappropriately. We need insight into factors which influence professionals’ own evaluations of genomic technologies, to guide effective and responsible implementation. A major threat to the utility and potential of genomic technologies is that of salience - perceived relevance to individual (and practices) is that of salience - perceived relevance to individual (and practices) is that of salience - perceived relevance to individual (and practices) is that of salience - perceived relevance to individual practice. We are conducting a pilot study of a method for engaging health professional audiences in evaluations of emerging genomic technologies which takes this into account. Objectives: To (a) understand how health professionals frame their understanding of emerging genomic technologies and (b) identify specific factors that may facilitate and challenge the appropriate adoption of these technologies in professional medical and nursing practice. Methods: A structured interactive workshop combines a largely didactic phase (devoting the majority of the time to the genomic technologies - (disease, risk, carrier status, medication response, genetic syndromes, metabolic disorders, newborn screening conditions, and ancestry)) and an interactive component which presents hypothetical case studies of different potential applications of genomics in routine practice. Participants engage in moderated discussion, and are encouraged to capture personal reactions and evaluation in un-analyzed notebooks. Participants are informed that their observations capture discussion points in real time. Survey data are captured before, during, and post-workshop. A mixed methods approach is taken for data analysis. Results: We have conducted workshops with physicians and nursing professionals in primary care and a range of specialty settings. Initial results summarizing participants’ evaluations of the utility of genomic approaches, and shifts in key attitudes compared with pre-workshop baseline, will be presented. The necessity of prompting participants’ evaluation of the salience of genomic interventions, as a prerequisite for recognizing the futility of research on specific applications, will be discussed. We will also discuss the challenge of developing a workshop format that is efficient in its use of participants’ time, and the relationship between professional engagement and continuing professional education exercises.
2509W

The awareness and need of genetic counseling service in Korea. H. Kim1,2
1) Medical Genetics, Ajou Univ Med Sch, Seoul, South Korea; 2) Chair, Genetic Counseling, Korean Foundation for Rare Disease(KFRD), Seoul, South Korea.

Genetic counseling service which is well recognized as an integral part of clinical genetics service deals with diagnosis and management of genetic conditions as well as genetic information presentation and family support in the developed countries of world. Korean Health care system is known for providing one of the most cost-effective services in the world. It is covered by the uniform national health insurance policy for which most people in Korea are mandatory policyholders. The necessity of genetic counseling service has been recently recognized by Korean medical communities, however, genetic counseling as an integral part of medical service is yet to be delivered to patients and their families in need. KFRD has held educational workshops and seminars on Genetic Counseling in eight university hospitals during 10 months period from Sep. 2011 to June 2012 to in order to educate and inform pts. and their families of rare disease, what is "genetic counseling service", and with full understanding of accurate "medical and genetic" information on the dis. involved, how can genetic counseling help pts. and their families to cope appropriately with consequence of the disease. Survey questionnaire were administered to 1663 attendees to evaluate the awareness and need of genetic counseling service. Among 1000 respondents, 79% reported that "previously never heard about genetic counseling service", 83% found that participation in workshop &/or seminar was helpful to comprehend the meaning of "genetic counseling" for them, and 98% indicated that "genetic counseling service" would be helpful to overcome with the condition for pts. and their families of rare disease. Based upon the result of the study, further educational workshops and seminars is required to provide further public awareness on genetic counseling. And it is clear that the need of genetic counseling service for pts. and their families of rare disease should be met as an integral part of medical services earliest possible for better management of dis. and quality of life. Further, more recognition and understanding of the fact that the scope and role of genetic counseling is expanding in post genomic era of personalized medicine for delivery of quality health care, will lead to the efforts to overcome obstacles in providing genetic counseling service in korean health care system.

2510W

Modelling of downstream counselling impact of ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. L. Burnett1,2,3, L. Deng1, D. Chesher1,2, R. Lew1,2, A. Proos1,2, 1) PaLMs Pathology North, NSW Health, Royal North Shore Hospital, St Leonards, Sydney, NSW, 2065, Australia; 2) Sydney Medical School - Northern, Royal North Shore Hospital E25, University of Sydney, Sydney, NSW, 2065, Australia; 3) Department of Obstetrics and Gynaecology, QEI Research Institute for Mothers and Infants, The University of Sydney, NSW, 2006 Australia.

Purpose: To predict clinical impact of expanded testing with Massively Parallel DNA Sequencing using the ACMG Recommendations for Clinical Reporting of Incidental Findings.

Design: Mathematical modelling of gene and variant frequencies.

Methods: A diagnostic panel of clinically significant genetic conditions was simulated, based on the American College of Medical Genetics and Genomics (ACMG) Recommendations. The number of patients with significant variants in one or more conditions that would be detected using the screening approach was calculated for different subsets of selected genes. Where a range of variant prevalence data was available, we selected the lowest and highest values, and calculated the most likely value as the geometric mean, while where only a single datum was available, we selected half and twice this prevalence as the lower and higher limits. Assuming all variants were simulated, based on the American College of Medical Genetics and Genomics (ACMG) recommendations, we calculated the rate of result type and age are retained [RR(path)=1.27, 95% CI (1.14,1.43); RR(Age)=1.42, 95% CI (1.25,1.60)]. In a two-factor model assessing the effects of result type and phenotype, the effect of pathogenic results is retained [RR=1.34, 95% CI (1.15,1.55)] but the effect of MCA is marginal [RR=1.36, 95% CI (1.02,1.82)]. In a two-factor model assessing the effects of result type and age, the effects of result type and age are retained [RR(path)=1.27, 95% CI (1.14,1.43); RR(Age)=1.42, 95% CI (1.25,1.60)]. Conclusions: Of the variables considered, phenotype and age are the strongest drivers of new medical recommendations in a pediatric cohort. Uncertain microarray results only drive medical management marginally. These findings further study with respect to establishing clinical utility.

Background Over 30% of colorectal cancer (CRC) is diagnosed at stage 3 due to low screening. Family history is the most important tool to identify hereditary cancer syndromes. Screening in families at hereditary risk reduces cancer morbidity and death. Prior studies demonstrate a low rate of family history documentation and low referral rates for genetic counseling and genetic testing. Aims 1) Assess the documentation of family history in different settings: hospital CRC versus office (breast cancer-BTWC); 2) Define barriers to family history documentation. Methods: We evaluated the documentation of family histories in charts of 630 consecutive colorectal cancer patients admitted for initial surgery in hospital and 295 Br Ca patients seen in office before their initial oncology consultations between 2009-2011. Statistical analysis was performed for each group of patients using normal method with standard error of proportion of 0.261 for 95% confidence interval (normal value 1.96) Of the 630 CRC patients, 237/630 (37%) had cancer family history recorded for the 1st degree relatives and 41/ 630 (7%) had cancer family history recorded for their 2nd degree relatives. The Amsterdam II criteria for referral to genetic cancer counseling and testing requires documentation of family history of 3 generations (first and second degree relatives and the proband): only 7% [95% CI 6.7%-7.7%] of the entire cohort of CRC patients, and only 6% of the colorectal cancer patients diagnosed under the age of 50 years old showed Amsterdam II criteria. In office: 213/295 (71%) [95% CI 47%-87%] had 3-generations pedigrees documented in their family history had documentation of family history and 82/295, (29%) did not have three generation pedigrees documented; Conclusions: Appropriate referral for genetic counseling and testing requires a complete and accurate documentation of family history. Significant differences were seen between the breast cancer charts and colorectal cancer charts, with greater accuracy of family history documentation and higher referral rates among breast cancer patients. Further improvements in the identification and management of patients at high risk and their family members, significant improvements in family history documentation are needed. Education is part of the answer.
2516W
Needs assessment of individuals with 22q11.2 Deletion Syndrome transitioning from pediatric to adult health care settings. W.L.A. Fung1,2,3,4, E. Leung1,3, A.S. Bassey1,2,3 1) Department of Psychiatry, University of Toronto Faculty of Medicine, Toronto, Ontario, Canada; 2) The Dalglish Family Hearts and Minds Clinic for 22q11.2 Deletion Syndrome, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada; 3) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 4) Departments of Psychiatry and Clinical Genetics, North York General Hospital, Toronto, Ontario, Canada; 5) Office of Continuing Education and Professional Development, University of Toronto Faculty of Medicine, Toronto, Ontario, Canada.
Background: The multi-systemic clinical manifestations of 22q11.2 deletion syndrome (22q11.2DS) make the provision of comprehensive care for patients challenging. A coordinated, multidisciplinary team approach in the provision of care is recommended in Practice Guidelines. While several comprehensive 22q11.2DS centers have been established for children - and the world's first such center for adults recently established in Toronto, Canada - the provision of such coordinated care had been identified as a particular challenge for 22q11.2DS patients transitioning from pediatric to adult health care setting. We sought to assess the needs of 22q11.2DS patients transitioning from pediatric to adult health care setting. Methods: A multi-pronged approach was utilized in assessing the needs of these patients. This included paper surveys of 22q11.2DS patients and families, phone interviews of 22q11.2DS patients and families, as well as electronic surveys of professionals involved in the care of these patients and families. Questions of both quantitative and qualitative nature were used. Results: Considerable differences were identified between 22q11.2DS patients and families in terms of their perspectives on the patients' 1) knowledge of 22q11.2DS; 2) ability to transition to adult health care setting and adult life overall; 3) social and emotional well-being. In general, patients rated themselves as having more knowledge regarding 22q11.2DS and greater social and emotional well-being, compared to their families' ratings. Patients also expressed a higher level of confidence in their ability to transition to adulthood than their families' rating. The key challenges identified by these families included psychological, emotional and behavioral issues associated with 22q11.2DS, the availability of social support to patients, and the negative effects of these symptoms on the patients' support systems. Findings from the professional stakeholder surveys will be presented at the annual meeting. Conclusion: To our knowledge, this is the first needs assessment conducted for 22q11.2DS patients transitioning from pediatric to adult health care setting utilizing a multi-pronged approach. These findings will help inform the development of an evidence-based transition program to enhance the care of these patients and their families.

2517W
The DSD-Translational Research Network, a national research and clinical network to improve health for people with Disorders of Sex Development. E. Delot1, D.E. Sandberg1, E. Vilain1. 1) Human Genetics, UCLA, David Geffen Sch Med, Los Angeles, CA; 2) Mott Children's Center, University of Michigan, Ann Arbor.
Disorders of Sex Development (DSD) are congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical. For families, the birth of a child with a DSD, with the uncertainty it brings regarding the child's gender and future psychosexual development, is believed to be extraordinarily stressful. Clinical care in DSD has been severely hampered by a fragmented research agenda, leaving fundamental gaps in knowledge of DSD pathology or treatment outcomes. Our long-term goal is to establish an environment in which clinical care of persons affected by DSD is evidence-based and guided by research that identifies factors impeding or enhancing opportunities for a positive quality of life across the lifespan. With the support of the NICHD, we created the DSD-TRN, to establish best practices by: 1) Building a sustainable infrastructure for translational research, including: - an interactive registry that will maintain standardized data from clinical sites, provide actionable information to clinical care teams and patients/families to support diagnosis and treatment management, and foster rapid translation of new evidence into clinical practice; - a collaborative network of researchers, clinicians, and patient/family advocates that will drive the research agenda and objectively monitor the impact of translational research at the bedside. 2) Standardizing radiological, biochemical, histological evaluations, descriptions of genital phenotype and post-surgical appearance and function. 3) Identifying biological and social factors associated with variability in psychosocial, psychosexual, and quality of life outcomes in patients with DSD. 4) Identifying novel pathophysiological mechanisms & improving the molecular diagnosis of DSD. The DSD-TRN network currently includes 5 academic centers (UCLA, U. Michigan, UCSF, Seattle Children's Hospital, Cincinnati Children's Hospital) with multidisciplinary teams (urology, genetics, endocrinology, psychosocial, etc.), multiple consultants (ethics, cancer, cost-effectiveness,...) and an Advisory Board with representatives of the major DSD patient advocacy groups. Deliverables already in place include a new exome-based genetic testing platform that allowed identification of novel variants in DSD patients, clinical use data collection forms integrated with electronic medical records, information brochures for patients and a registry of patient data to support future evidence-based clinical practice.
Whole genome sequencing vs. family history: physician perceptions of clinical utility, J.L. Vassy1,2,3, R.C. Green1,2,3, J. Kher2, D. Lautenbach2, K.D. Christensen4, M.A. Giovanni2, M.F. Murray2, A.L. McGuire5 for The MedSeq Project. 1) Department of Medicine, Harvard Medical School, Boston, MA, USA; 2) Division of General Medicine and Primary Care, Brigham and Women’s Hospital, Boston, MA, USA; 3) Section of General Internal Medicine, VA Boston Healthcare System, Boston, MA, USA; 4) Baylor College of Medicine for Ethics and Health Policy, Houston, TX, USA; 5) Medical Genetics Training Program, Harvard Medical School, Boston, MA, USA; 6) Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA; 7) Geisinger Health System, Danville, PA, USA.

Background The uptake of whole genome sequencing (WGS) into patient care will depend in large part on whether physicians perceive it to have clinical utility. It is informative to compare the perceived utility of WGS to the current benchmark of genetic risk assessment in clinical care: family history (FmHx). Standard medical practice includes at least a cursory FmHx assessment, while for most physicians WGS remains an experimental technology. We hypothesized that physicians would find FmHx more clinically useful than WGS now but that they expect the utility of WGS to increase in the future. Methods We have enrolled primary care physicians and cardiologists to participate with their patients in the MedSeq Project: a randomized trial of WGS in clinical care. Before enrolling patients, we surveyed physicians about their perceived clinical utility for FmHx and WGS. We asked them to rate on a scale of 1-10 (‘Not at all Useful’ to ‘Extremely Useful’) how useful they thought the study’s FmHx and WGS reports would be for managing their patients’ health at two times: now and in the future. We categorized responses as not useful (1-5) or useful (6-10) for descriptive analyses. We used paired t-tests to compare the reported utility ratings of FmHx vs. WGS. Results Of 17 physicians (mean age 52 years), 9 (53%) were women, and 5 (29%) were of non-white race. Nine (53%) were primary care physicians, and 11 (65%) reported no genetics training beyond the usual medical school curriculum. The majority of physicians said FmHx would have clinical utility both now and in the future (n=15, 88%, for both responses). In contrast, only 4 (24%) said that WGS would have utility now. However, 15 (88%) predicted that WGS would have future utility for their patients’ care. Physicians rated FmHx to have higher utility than WGS now (mean responses 7.6 vs. 4.2, p<0.001), but they rated FmHx and WGS to have similar utility for the future (mean responses 7.9 vs. 7.5, p=0.49). Conclusions Physicians participating in the MedSeq Project rated FmHx to have greater utility than WGS for the present-day management of their patients’ health, but they expected this utility gap to close in the future. Whether and how quickly this gap closes will likely depend on how the clinical genomics community addresses barriers to integrating WGS into patient care.
Genetic counsellors' preferences for Preimplantation Genetic Diagnosis: Designing a Discrete Choice Experiment. E. Goh1, 2, W. Ungar2, D. Marshall2, P. Miller1. 1) Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; 2) Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada; 3) Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada.

Preimplantation genetic diagnosis (PGD) permits couples at high risk of a genetic condition to test an embryo for it prior to pregnancy. In the absence of explicit Canadian public policy regarding conditions of use, the decisions of front line clinicians (genetic counsellors, GC) guide clinical practice.

Discrete choice experiments (DCE) are a valid method for quantifying preferences and for measuring trade-offs between the characteristics of alternatives. Our qualitative research included 2 focus groups of practicing GC in Toronto, using a semi-structured focus group guide to identify factors that are important to GC in recommending public PGD coverage to inform our DCE. The first focus group explored what GC deemed relevant using open probes. These criteria were combined with a literature review as probes in the second focus group to generate a list of common coverage criteria. GC ranked the criteria individually, drawing on their perception of importance. Field notes and rankings were descriptively analyzed to identify insights surrounding coverage criteria. Pre-testing of the draft DCE survey was completed using a cognitive interview protocol with 7 GC.

The following attributes were identified as relevant for inclusion: i) PGD indication (childhood-onset condition, adult-onset condition, and adult-onset cancer predisposition), ii) risk of the genetic condition (50%, 25% and 1%), iii) fertility status of couple (infertile or fertile), iv) family history (have children at least one of who is affected, have children none of whom are affected, or have no children) and v) number of in vitro fertilization cycles to be funded (1 cycle, 3 cycles or 6 cycles). Based on pre-testing, PGD indication appeared to be the most important factor in deciding between the alternatives.

This qualitative study identified factors deemed important to GC in informing public PGD coverage. Most, but not all of the factors were highlighted in the literature. A DCE was designed based on these attributes. A pilot study of the quantitative phase is currently being undertaken (results expected July 2013, N=20). The results of the DCE can be used to quantify trade-offs amongst these factors and quantify national GC preferences for PGD, which can help inform public policy.


Statement of purpose. By 2004, genetic consultations, molecular diagnosis and research in human genetics in Peru were performed in a few centers in Lima, the country's capital city. In the last decade, genetic services have significantly grown but are still unregulated. Our study aims to identify and characterize the institutions offering human genetic services in Peru.

Methods. A cross-sectional study was carried out using web search and interviews that were conducted with geneticists and other field-related professionals from institutions offering clinical, molecular and/or research services in human genetics in Peru. The survey was realized using a standard collecting data form. The study obtained IRB approval.

Results. From the sixty-two identified institutions, forty of them completed the survey. The majority (92.5%) are located in Lima. Public health institutions represent 25%; five percent are public forensic institutions; 10% belong to both private and public universities, from which only two offer postgraduate training in human genetics; 5% are non-profit organizations; 55% are private genetic centers dedicated to specific laboratory activities including karyotyping (52.5%), FISH (32.5%), paternity testing (17.5%) and other molecular diagnostic techniques, such as PCR (50%), qPCR (27.5%), gene sequencing (27.5%) and others (25%). More than half of the participating institutions (55%) claim to perform basic and clinical research; only ten institutions have published a total of thirty-eight MEDLINE indexed papers. Twenty-one institutions offer genetic counselling and four some kind of genetic treatment. Conclusions. Despite scarce training opportunities, human genetic services are growing in Peru but are still mainly located in Lima. We envision this information to be helpful in generating human genetic networks and defining public health policies in Peru.
Diagnostic application of targeted resequencing for familial nonsyndromic hearing loss. B. Choi1, G. Park2, J. Kim2, A.R Kim2, B.J Kim2, T. Park3, S.Oh2, K. Han2, W. Park2,4 1) Seoul National University, Seoul, South Korea; 2) Sungkyunkwan University, Suwon, South Korea; 4) Samsung Genome Institute, Samsung Medical Center, Seoul, South Korea.

Identification of causative genes for hereditary nonsyndromic hearing loss (NSHL) is important to decide treatment modalities and to counsel the patients. Due to the genetic heterogeneity in sensorineural genetic disorders, the high-throughput method can be adapted for the efficient diagnosis. To this end, we designed a new diagnostic pipeline to screen all the reported candidate genes for NSHL. For validation of the diagnostic pipeline, we focused upon familial NSHL cases that are most likely to be genetic, rather than to be infectious or environmental. Among the 32 familial NSHL cases, we were able to make a molecular genetic diagnosis from 12 probands (37.5%) in the first stage by their clinical features, characteristic inheritance pattern and further candidate gene sequencing of GJB2, SLC26A4, POU3F4 or mitochondrial DNA. Next we applied targeted resequencing on 80 NSHL genes in the remaining 20 probands. Each proband carried 4.8 variants that were not synonymous and had the occurring frequency of less than three among the 20 probands. These variants were then filtered out with the inheritance pattern of the family, allele frequency in normal hearing 80 control subjects, clinical features. Finally NSHL-causing candidate mutations were identified in 13 (65%) of the 20 probands of multiplex families, bringing the total solve rate (or detection rate) in our familial cases to be 78.1% (25/32). Damaging mutations discovered by the targeted resequencing were distributed in nine genes such as WFS1, COCH, EYA4, MYO6, GJB3, COL11A2, OTOF, STRC and MYO3A. We developed a new software tool at least for familial NSHL to find mutations based upon its efficacy and cost-effectiveness.

Automating Clinical Exome Analysis. M.N. Bainbridge1,2,3, E.B. Venner2, C. Eng2, Y. Yang2, R.A. Gibbs1,2 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Whole Genome Sequencing Lab, Baylor College of Medicine, Houston, TX; 3) Codified Genomics, LLC, Houston, TX.

The clinical use of genome wide sequencing is now commonly available and is poised to be ubiquitous in the next year, however, several challenges remain in the timely and cost effective analysis of these data. Initial review and prioritization of the hundreds of potentially disease causing variants is critical for making accurate diagnoses. Unfortunately, this first step is expensive, laborious, time consuming and prone to error. We have developed an information pipeline and set of algorithms to fully automate this initial review process. Initially, we capture clinician expertise regarding the patient phenotype and use this information to prioritize genes based on phenotypic overlap with diseases. This enable those with less medical training to identify variants which likely contribute to the patients phenotype and promote these variants for further review. Additionally, using this data as well as a rich set of annotation information about each variant and gene, we can automate the categorization of the variants according to ACMG guidelines. We tested our algorithms on a series of clinical exome cases that were previously solved by standard methods and found that the causative variant was almost always within the top 15 of all rare, protein changing variants. Further, we were able to identify other, strong candidate variants that were not highly prioritized using manual methods including in cases where multiple genes were suspected of contributing to the phenotype. Further, these improvements reduce the time spent on the initial variant review to approximately half. These analysis methods are critical in a clinical setting. They aid in reducing costs, time, and errors in the initial review greatly improving the quality of the test.


Horizon Discovery has established a range of best in class, genetically defined, genomic reference standards, including FFPE blocks and purified gDNA. These standards offer a sustainable and highly defined source of reference material to laboratories, proficiency schemes, and manufacturers. Here we present the design and validation of our Next Generation Sequenc- ing (NGS) quantitative multiplex reference standards. NGS offers significant advantages for mutation detection, enabling the simultaneous detection of multiple mutations in multiple genes, and provides a digital readout of the mutation frequency. Challenges remain however, not least the number of different platforms, each presenting different systematic bias, and the need for extensive validation of the analytical pipeline to ensure variant calling is correctly and consistently achieved. A method for standardisation and verification is therefore required that goes beyond the current widespread use of HapMap controls. In response to the need for a better NGS reference standard, Horizon has leveraged its proprietary genome editing technology to create a multiplex reference standard covering many commonly assayed cancer mutations. The power of this approach is that virtually every character- istic of our reference standard can be defined, from the molecular constitu- tion of the genome, to the DNA output associated with each product batch. Each sample contains key oncogenic mutations including KRAS G13D and G12D, PIK3CA H1047R, BRAF V600E, EGFR T790M, NRAS Q61K and KIT D816V represented at defined allelic frequencies ranging from 1 to 25%. Furthermore, >20 defined mutations in other disease relevant genes including RAS, CTNNB1, ALK, FGFR2, MET, IDH1, NOTCH and BRCA1 are present. Digital PCR is used to quantify each mutation, enabling the precise quantification of mutational frequency. We show that the data obtained by NGS on our multiplex standard closely matches our digital PCR data, thus validating the use of the standard as a benchmark for reference material by any lab intending to use NGS for clinical diagnostics. The multiplex reference standard described here establishes the integrity of the NGS workflow from enrichment and sequencing through to bioinformatics and data interpretation.

Simultaneous detection of point mutations and exonic deletions by target gene capture and deep sequencing. Y. Feng, G.L. Wang, H. Cui, J. Wang, Y.W. Zhang, L.J. Wong. Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Background: Next generation sequencing (NGS) has demonstrated its clinical utility in the identification of point mutations but not the detection of large intragenic deletions, which is extremely important in elucidating its presence as a compound heterozygous mutation to a point mutation or accounting for an apparently homozygous mutation in autosomal recessive cases. Method: An analytical algorithm to detect exonic deletions was developed by comparing the normalized coverage depth (CD) of each coding exon of the testing sample to CD of that particular exon from a group of controls. Results: This algorithm allows detection and confirmation of homozygous and heterozygous deletions of various captured genes. This includes homozygous deletions of MPV17 exons 3-7 deletion of a patient with mtDNA depletion syndrome, LPINT1 exon 18 of a patient with rhabdomyo- lysis, and a hemizygous PHKA2 exons 27-30 deletion of a male patient with GSD. In addition, there are compound heterozygous exonic deletions with point mutations, including CPST E9-11, citrin E3, TK2 E1-2, partial OTC exon 2 deletions, and a CACT E5-9 deletion originally identified as an apparently homozygous point mutation. Conclusion: It is extremely valuable to be able to detect point mutations and large exonic deletions simultane- ously using target gene capture/NGS approach.
2526T
Next Generation Sequencing coupled with a Novel Multiplex PCR Protocol for Comprehensive Genetic Screening of Maturity Onset Diabetes of the Young in India. A. Chapla1, D.M. Mahesh1, D. Varghese2, S.V. Nadig1, H.S. Asha1, R.T. Varghese3, M. Inbakumar1, F. Christina1, S. Matha2, T.V. Paul1, N. Thomas1. 1) Department of Endocrinology, Diabetes and Metabolism, Christian Medical College, Vellore, Tamil Nadu, India; 2) Department of Pediatric Endocrinology, Christian Medical College, Vellore, Tamil Nadu, India.

Maturity Onset Diabetes of the Young (MODY) is a monogenic disorder with an autosomal dominant pattern of inheritance characterized by β-Cell dysfunction and accounts for around 1-2% of type 2 diabetes and typically presents before the age of 25 years. Till date mutations in at least 13 different genes have been reported to cause MODY. Due to an overlap of clinical features with polygenic Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D), identification of patients with MODY is a diagnostic challenge. MODY genetic screening is of immense clinical importance and a confirmed genetic diagnosis would help streamline therapy. However, due to limitations in the scalability of the current diagnostic platform, performing genetic screening of a comprehensive panel of the identified MODY genes has been hindered in the past. As a result, majority of MODY patients are stereotypically classified as T1D or T2D and may potentially receive inappropriate therapy. This study aimed to establish genetic screening of a comprehensive panel of 10 MODY genes consisting of HNF1A, HNF4A, GCK, HNF1B, IPF1, NEUROD1, KLF11, CEL, PAX4, and INS. A novel multiplex PCR was established to enrich the target genes and further sequencing was performed on Ion Torrent Personal Genome Machine. Using this approach 95% of the targeted DNA was covered at a depth of 50 or more reads per nucleotide and at a minimum average coverage depth of 20X. MODY genetic testing was carried out in 50 subjects with young onset diabetes of which 35 met the clinical criteria of MODY. We identified mutations in seven patients, which include four with NEUROD1 mutation (novel c.175 G>C p.E59Q, novel c.182G>A SUTFR, and c.193C>G p.H241Q), one with HNF4A mutation (c.505G>A p.V169I cosegregating with c.493-4G>A and c.493-20C>T), one with GCK mutation (c.1318G>T p.E440X) and one with HNF1B mutation (novel c.274 C>T p.L92F). These mutations and the other identified rare variants were confirmed by Sanger sequencing. Further, to validate the protocol we compared the Ion torrent sequencing data of HNF1A with conventional Sanger sequencing. Further, we have identified a higher frequency of NeuroD1 mutations, a pattern of MODY different from the Western population.

2526W
Next generation molecular diagnosis of patients with retinal degeneration. L. Lan, N. Li, J. Chiang. Casey Molecular Diagnostics Laboratory, Portland, Oregon.

Molecular diagnosis of inherited retinal degeneration has been difficult due to the following reasons: (1) locus heterogeneity (more than 200 genes involved); (2) allele heterogeneity (many private mutations); (3) overlapping clinical presentations; and (4) progressive nature of some conditions. The arrival of mass parallel sequencing: Next Generation Sequencing (NGS) provides an opportunity to revolutionize the task. At the Casey Molecular Diagnostic Laboratory, we have developed a method combining the specificity and low cost of PCR enrichment with Illumina MiSeq NGS platform. Several hundred clinical samples have been sequenced using this combination of methods thus far. Based on our experience, false positive and false negative callings are concerns. The presence of low coverage regions (gaps) is another concern. Mutations can be missed below cutoff (false negative), while sequencing artifacts above cutoff generate false positive results. Homozygous deletions can be missed if gaps are not filled. Therefore, after data analysis, gaps are also filled in by PCR and Sanger sequencing, and mutations and novel variations are confirmed by Sanger sequencing regarding the quality score. The biggest drawback of this individual disease panel approach is for patients with uncertain clinical diagnoses, as multiple panel testing may be required in order to find mutations. At this time, we are developing a PCR based approach of testing the entire non-syndromic retinal degeneration genes plus some common syndromic genes. With the buildup of our database by testing more genes and confirming mutations and novel variations by Sanger sequencing, we aim to improve molecular diagnoses of patients with retinal degeneration. Our unique approach and experience will be presented.

2527F
Many types of DNA damage can be detected with Two-Dimensional Strandness-Dependent Electrophoresis (2D-SED). J. J. Jonsson1, 2, B. Gudmundsson1, 2, H.G. Thormar1, A.G. Sigurdsson1, S. Thongthip1, M. Steinardottir1, A. Smorogzewska1, 3, 4, A. Mathai1, 4, 5, T.V. Paul1, N. Thomas1. 1) Department of Genetics and Molecular Medicine, Landskapi, Reykjavik, Iceland; 2) Dept. of Biochemistry and Molecular Biology, University of Iceland, Reykjavik, Iceland; 3) Lifefone/BioCure Inc., Reykjavik, Iceland; 4) Laboratory of Genome Maintenance, Rockefeller University, New York, NY.

Two-Dimensional Strandness-Dependent Electrophoresis (2D-SED) in manual minigels or premade microgels is a novel technique for nucleic acid analysis. In the first dimension nucleic acid fragments are separated based on length and strandness i.e. double-stranded DNA, single-stranded DNA and RNA-DNA hybrids. The nucleic acids are heat denatured before the second dimension electrophoresis and in the second dimension all fragments are single-stranded and separated only based on length. We tested if 2D-SED could detect various types of DNA damage in vitro and in vivo. Each sample was run in duplicate both uncut and cut with Mbo I, an enzyme which cuts both single- and double-stranded DNA. Single-stranded breaks, either nicks or gaps, were detected as horizontal streaks on 2D-SED extending from uncut DNA molecules too large to efficiently enter the gel. Double-stranded breaks of depth 20X, 10X and 5X were detected with double stranded DNA molecules and single stranded DNA molecules with interstrand crosslinks (ICL) migrated as an arc behind normal dsDNA molecules. In contrast, DNA with intrastrand crosslinks and bulky adducts were bent and migrated in front of that arc. Single-stranded DNA molecules, too damaged for complementary strand binding, formed a diagonal line. 2D-SED detected DNA damage at comparable level of sensitivity to the well known comet assay. However, 2D-SED allowed direct detection of damaged molecules and subfractions could be isolated form the gel. 2D-SED could also be used to analyze 2D-SED can detect single types of DNA damage and the effects of DNA repair. Applications include testing quality of biosamples and efficiency of various molecular procedures were damage to DNA is common. Examples include extensive damage detected in CHIP-Seq experiment, TFFPE samples. Applications also include genotoxicity testing, chemosensitivity testing and diagnosis of genome instability and DNA repair disorders.

2527T

Newborns are disproportionately affected by rare monogenic diseases including hearing loss, inborn errors of metabolism, and lysosomal storage disorders. A diagnosis is critical in order to begin treatment early and prevent mortality or lifelong debilitation. Unfortunately, the current tests for newborns are limited in scope and scalability. Diagnostic confirmation often includes time-consuming and expensive serial single gene testing, which can result in reducing the available treatment window and increasing the risk of long term effects. We believe that next generation sequencing technology has the power to help infants in a way that conventional diagnostic approaches cannot. Targeted next-generation sequencing (TNGS) has allowed expansion of genetic testing across a large number of diseases. Here, we report on a new multi-gene panel that targets hundreds of genetic diseases predominantly affecting newborns and which neonatologists have recommend for inclusion in population based screening. Our panel encompasses diseases and symptoms such as hearing loss, hypotonia, hepatomegaly and failure to thrive. Through our TGNS pipeline, sequencing capacity can be maximized for high coverage of targeted regions along with decreased turnaround time and increased throughput. This optimized pipeline with lower per sample cost and better accuracy in variant calling can translate in the clinic to an accelerated patient diagnosis. Our TGNS approach focuses on the exonic regions of approximately 300 genes and uniquely also includes the entirety of specific genes, such as those for hearing loss and cystic fibrosis, where variants of clinical relevance are often found in non-coding regions. This design expands detection beyond exclusively exonic panels while maintaining the advantages of a smaller targeted panel. In our lab we additionally demonstrate improvements to extraction methods for small sized samples from newborns (e.g. DBS, saliva) in order to yield DNA in quantities sufficient for our TGNS approach by hybrid-capture and sequencing. Thus mitigating the need for whole genome amplification and associated biases. Our pipeline provides a sample to answer solution by incorporating our advances in sample prep, reduced sequencing run times with the latest instruments, and implementing commercial grade bioinformatic analysis with automated variant calling tools and interfaces.
Targeted Sequencing of Genes Causing atypical Hemolytic Uremic Syndrome and Coagulation Disorders. S. Theru Arumugam, K. Meghana-than, D. Kissell, S. Jodele, R. Grupp, K. Zhang. 1) Division of Human Genetics, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH; 2) Division of Hematology, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH; 3) Division of Bone Marrow Transplantation and Immune Deficiency, Cincinnati Children’s Hospital Medical Center, Cincinnati, Ohio.

Sanger sequencing of multigenic disorders can be technically challenging, time consuming and very expensive. Recent advancements in Next-generation sequencing (NGS) not only facilitated the discovery of new disease genes, but also transformed the routine clinical diagnosis of genetic diseases. In this study, we evaluated the performance of NGS in detecting the DNA sequence variations in 52 genes causing atypical hemolytic uremic syndrome and coagulation disorders. Targets covering the exons, flanking intronic and regulatory regions of these genes were enriched using Rain-Dance microdroplet PCR and sequenced, as single end 50 bp reads, on the Illumina HiSeq 2500 instrument. The raw sequence reads from each sample were subjected to molecular genetics laboratory’s (MGL) NGS data analyses procedure, which include: (1) quality filtering of raw reads (2) mapping high quality reads to reference sequence and simultaneously detecting sequence variants using NEXTGene software (3) identifying potential true sequence variants using MGL criteria and (4) confirmation of variants with Sanger sequencing. A total of eight samples with known mutation were included in the pilot study and an average of about 20 million raw sequence reads were generated from each sample. About 97% of reads in each sample met the quality criteria and of which, approximately 67.5% aligned to the reference sequences of these 52 genes. We observed that about 98.5% of > 145 kb target regions, covering exon and 20 bp flanking intronic regions, had at least 40 sequence reads. The data is currently being analyzed and results of this study will be presented.


Purpose: The challenge in genetic diagnostics is to apply one comprehensive test for heterogeneous diseases. The number of genes to examine in a particular clinical case can be relatively large because phenotypes of many hereditary syndromes are known to overlap and different genes may underlie a single syndrome. Our aim was to design and implement various targeted next generation sequencing (NGS) gene-panels, starting with those disorders that account for the majority of current diagnostic requests. Methods: As proof of principle we developed a gene-panel based on Agilent Sure Select Target Enrichment® for simultaneous mutation detection for 48 genes associated with hereditary cardiomyopathies. To assess test-sensitivity and specificity we performed a validation on 84 patients. For 24 of these Sanger Sequencing (SS) data for up to six genes were available. Pools of 12 samples were sequenced using 151 bp paired-end reads on an Illumina MiSeq® sequencer and analyzed using NextGene and Cartagenia-NGS® software. Subsequently, an extended gene-panel targeting 55 cardiomyopathy-associated genes was implemented in routine diagnostics. To date, more than 200 patients have been analyzed. In parallel, a targeted NGS gene-panel for 70 cancer predisposition genes was validated in a similar manner. Results: For the cardiomyopathy gene-panel, 99 percent of all bases had a coverage of ≥30 reads per nucleotide. Because of poor coverage a total of 11 regions were analyzed in parallel using SS. We identified ≥21000 variants (≥250 per patient). SS was performed for 168 variants (155 substitutions, 13 indels). All were confirmed, including a deletion of 18 bp and an insertion of 6 bp. Compared to previous routine diagnostics based on SS, application of our NGS-method resulted in increase of diagnostic yield from 15 percent to about 50 percent. Results for validation of the cancer gene-panel were comparable: 99 percent of all bases with ≥30 reads per nucleotide, for 22 regions SS is performed in parallel, no false positive or negative results upon confirmation with SS of 180 variants. The first patients are now being tested in diagnostics. Conclusion: Targeted NGS of a disease-specific subset of genes can be reliably implemented in diagnostics to analyze large numbers of genes in parallel with significantly improved diagnostic yield. Additional gene-panels i.e. for epilepsy and neurodegenerative disorders are currently being designed and implemented.


The main challenge for clinical targeted next-generation sequencing methods is obtaining complete and uniform coverage of all target regions. Some popular methods for target enrichment rely on lengthy and inefficient hybrid capture or multiplexed PCR techniques, resulting in lower coverage and more off target reads. To address these challenges, WaferGen has developed the high-density SmartChip TE System. One SmartChip TE panel supports conducting hundreds to thousands of parallel, singleplex PCR reactions to efficiently enrich desired target regions in less than 3 hours. Results of a recent study used a SmartChip TE Panel (140 kb target region) targeting exonic regions in a 17 gene set in 16 cancer cell lines (NCI 60). High design rates and percent bases on target ensured specific amplification necessary for efficient enrichment. The sequencing results from SmartChip TE enriched targets showed coverage at 20x and 100x of 98.8% and 98.2%, respectively with a uniformity of coverage of 98.8% ± 10% of mean coverage. Comparison of performance between AmpliSeq and HaloPlex designs and the SmartChip TE custom designs will be presented. The unprecedented enrichment quality is achieved with flexibility of running up to 2500 unique singleplex reactions on a single SmartChip TE chip. The results indicate that CLIA-certified or clinical research laboratories can utilize the SmartChip TE system to obtain the most complete and uniform coverage among target enrichment technologies.


Current genetic tests for diagnosing monogenic diabetes rely on selection of the appropriate gene for analysis according to the patient’s phenotype. Next generation sequencing enables the simultaneous analysis of multiple genes in a single test. Our aim was to develop a targeted next generation sequencing assay to detect mutations in all known MODY and neonatal diabetes genes.

We selected 29 genes in which mutations have been reported to cause neonatal diabetes, maturity-onset diabetes of the young (MODY), maternally inherited diabetes and deafness (MIDD) or partial lipodystrophy (FPLD). We designed an exon-capture assay to include the coding regions and conserved splice sites. A total of 114 patient DNA samples were tested: 32 with known mutations and 82 previously tested for MODY (n=33) or neonatal diabetes (n=49) but in whom a mutation had not been found. Sequence data were analysed for the presence of base substitutions, small indels and exonic deletions or duplications.

All known mutations and polymorphisms (n=70 different variants) were detected including 55 base substitutions, 10 small insertions or deletions and 5 partial/whole gene deletions/duplications. Previously unidentified mutations were found in 5 patients with MODY (15%) and 9 with neonatal diabetes (18%). Most of these patients (12/14) had mutations in genes that had not previously been tested. These included mitochondrial r.3243A>G mutation; previously missed intronic HNF4A (c.358+5G>A) mutation; mutations in EIF2AK3 and SLC19A2 not originally tested because patients were referred before the onset of syndromic features; ABCG8 mutation that was missed because of allelic dropout caused by a polymorphism within a primer binding site; and novel mutations in GCK, PDX1 and GATA6.

In conclusion, our novel targeted next generation sequencing assay provided a sensitive method for simultaneous testing of all monogenic diabetes genes. The increased number of genes tested led to an improved mutation detection rate.
NGS Data Analysis for a Primary Immunodeficiency Gene Panel using Haloplex Enrichment Method. J. Durttschi, E.M. Coonrod, R.L. Margraf, H.R. Hill, K.V. Voelkerding, A. Kumanovic, ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 2) Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT; 3) Department of Medicine, University of Utah School of Medicine, Salt Lake City, UT; 4) Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT.

Haloplex (Agilent) gene target enrichment offers a cost effective and efficient workflow for library preparation followed by illumina MiSeq sequencing and was used to develop a panel targeting 94 primary immunodeficiency associated genes. Haloplex capture uses restriction enzymes to digest genomic DNA and then probes selectively capture and amplify DNA in 100 to 500 base DNA segments tiled across target regions. Haloplex sequence data is characterized by non-randomly distributed sequence inserts defined by the limited set of restriction enzymes used and unwanted illumina adapter sequence at read 3’ ends when sequencing goes beyond the end of shorter Haloplex insert sizes. In this study we evaluate Haloplex data processing and analysis issues encountered in our 94 gene panel. A custom Haloplex panel was designed for 94 genes associated with primary immune deficiency disorders and used to prepare NGS libraries for 40 samples. Four samples were indexed and multiplexed per illumina MiSeq, 2x150, paired end run. Read data was trimmed of MiSeq adaptor sequence present at some 3’ read ends, aligned, and analyzed using Cutadapt, BWA, and GATK software, respectively. During Haloplex data processing, illumina adapter sequence was identified and trimmed from an average 40% of reads amounting to 13% of bases trimmed per data set. After this trimming, over 10% of reads were shorter than 100bp suggesting that these reads came from unintentional Haloplex inserts below the 100bp minimum. Inferred insert size analysis of alignments also indicated insert sizes less than 100bp for over 10% of read pairs. Median coverage in coding regions of panel genes was much higher in Haloplex samples versus three typical Nimblegen V3 exomes (on average, 1062 versus 123) but the fraction of the same coding regions with coverage below 30 was higher in Haloplex versus exomes (11% versus 4%) indicating less uniform coverage of Haloplex data. Of the collection of 702 different variant classes versus 512 in Nimblegen V3 exomes, 526 (75%) were present in at least one of 40 samples, 23% were not seen in the 5,400 exome data set, indicative of a high false positive rate. However, many likely false positives appear at systematic, trackable locations. Our ongoing study indicates that, when properly managed, these Haloplex data characteristics lead to an effective enrichment method for our multi-genene panel assay for the clinical testing of primary immunodeficiencies.
2537W
Next-generation DNA sequencing (NGS) is poised to displace genotyping technology for clinical applications because it promises richer information at low cost. However, to date, the number of consensus guidelines including sequence accuracy and completeness, as well as workflow scalability, have limited its adoption in the clinical laboratory. Here we describe a NGS-based platform designed for genetic carrier screening in a clinical setting. A set of 15 genes are isolated from genomic DNA by automated multiplex target capture tagged with molecular barcodes, and pooled and sequenced on the Illumina HiSeq system. Reads from each sample are de-multiplexed, aligned to a reference, and integrated into accurate genotype calls which are then interrogated for pathogenic mutations. A total of 42,856 bp were targeted for capture by a set of molecular inversion probes designed to lie across the target such that each base was captured by at least three different probes. Across a set of 182 DNA samples derived from cell lines or blood, a median of 99.71% of bases were sequenced to sufficient depth and quality for inclusion in genotype calling. Repeatable runs exhibited a high level of concordance, with 17 discordant single nucleotide variant (SNV) calls out of 5,177,206 across 126 samples. Concordance of genotype calls with bidirectional Sanger sequencing of PCR amplicons derived from a set of 194 samples was 99.97% at SNV positions (1 out of 4,001 Sanger SNV calls discordant), and 99.999% at non-variant positions (8 out of 6,992,754 Sanger non-variant calls discordant). We identified, in a set of 55 samples, a total of 92 mutations (out of 92) that have been previously reported to be causal in known Mendelian disease (MDD) and 14 previously uncharacterized samples contained mutations that were either known or expected to be pathogenic. Clinical carrier screening has traditionally been performed using genotyping technology, and demands high analytical accuracy and turnaround time. Targeted NGS platforms with exceedingly high concordance with Sanger sequencing, long considered an accuracy gold standard. Furthermore, we made high-confidence genotype calls across the vast majority of basepairs within our target regions. Taken together, these results indicate that NGS, when paired with the appropriate sample preparation methodology, automation, and data analysis, can deliver the performance required for use in clinical carrier screening.

2538T
Targeted exome sequencing identifies two pathogenic DYNC2H1 variants in a fetus with short-rib-polydactyly syndrome. K.I. Varvaglia-anis,1,2, P. Makrythanasis,2, F. Santon2, J.-M. Pellegrini2, P. Externmann, C. Brockmann1, C. Gehrig1, M. Giupponi1,3,4, J.-L. Bouin1,3,4, S.E. Antonarakis1,2, S. Fokstuen. 1) University Hospitals of Geneva, Service of Genetic Medicine, Genomic Clinic, Geneva, Switzerland; 2) University of Geneva, Department of Genetic Medicine and Development, Geneva, Switzerland; 3) University Hospitals of Geneva, Department of Gynecology and Obstetrics, Geneva, Switzerland; 4) Dianecho, Geneva, Switzerland.
The short-rib-polydactyly syndromes (SRPS) represent a heterogeneous group of ciliary skeletal dysplasias characterized by a narrow thorax, short ribs and limbs and polydactyly. To date, more than 10 genes have been associated with this mainly autosomal recessively inherited phenotype.

We applied targeted exome sequencing on an affected fetus presenting with micromelia, postaxial polydactyly of both hands and feet, short ribs and a severe cardio-thoracic disproportion. The pregnancy was terminated at 16+3 gestation weeks based on the ultrasonographic finding of a severe cardio-thoracic disproportion. In the postmortem pathology, we determined that 5,177,206 across 126 samples. Concordance of genotype calls with bi-directional Sanger sequencing of PCR amplicons derived from a set of 194 samples was 99.97% at SNV positions (1 out of 4,001 Sanger SNV calls discordant), and 99.999% at non-variant positions (8 out of 6,992,754 Sanger non-variant calls discordant). We identified, in a set of 55 samples, a total of 92 mutations (out of 92) that have been previously reported to be causal in known Mendelian disease (MDD) and 14 previously uncharacterized samples contained mutations that were either known or expected to be pathogenic. Clinical carrier screening has traditionally been performed using genotyping technology, and demands high analytical accuracy and turnaround time. Targeted NGS platforms with exceedingly high concordance with Sanger sequencing, long considered an accuracy gold standard. Furthermore, we made high-confidence genotype calls across the vast majority of basepairs within our target regions. Taken together, these results indicate that NGS, when paired with the appropriate sample preparation methodology, automation, and data analysis, can deliver the performance required for use in clinical carrier screening.

2539F
Validation of an Accurate, High-Throughput Multiplex qPCR Assay to Confirm CMA Clinical Findings. L.E. Northrop, V. Aggarwal, V. Jobanputra, M. Mansukhani, B. Levy. Department of Pathology & Cell Biology, Division of Personalized Genomic Medicine, New York, NY.
Chromosomal microarray analysis (CMA) is the standard cytogenetic screening method in clinical constitutional laboratories, being used in patients with developmental delay and intellectual disability. The demand for CMA in perinatal patients is likely to increase following the findings of the recent NICHD multicenter prenatal study, recognizing the need for a more comprehensive approach to fetal karyotyping. With the recent advances in NGS technology, automation, and data analysis, can deliver the performance required for use in clinical carrier screening.

2540W
Detection and quantification of somatic mutations in Klippel-Trenaunay Syndrome using digitally counted nanodroplets. N.M.K. Kamitaki1, V. Lucks2, R. Murillo3, S.A. McCarrory, M. Warman1,2,3, M. Reinhold4, S. Sagel5, S. Schill6, J. Niazi7, C. Demeur8, A. Kamitaki1, A. Saifi1, L. Calvo1, C. Webster1, B. Levy. 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA; 3) Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA.
All human tissues contain cells with somatically acquired mutations. Those mutations that cause disease may be difficult to identify. When the cells of an embryo having a somatic mutation is high in an available tissue, the disease causing mutation can be identified using commonly applied sequencing methods such as massively parallel sequencing of exomes or genomes, or PCR-based amplifications and sequencing of candidate genes. When the frequency of mutant cells is low, mutation detection becomes much more challenging with these methods. Our molecular approach is based on using allele-specific digital PCR in nanodroplets. We designed primers and allele-specific probes, where each reference and mutation-specific probe contains the materials requisite for PCR, but only some of which contain a template DNA molecule from the patient’s genome. We count the number of droplets positive for each fluorophore using a Bio-Rad QX1000 Droplet Reader. This allows us to do high-throughput counting of the abundance of each allele, allowing us to not only detect if mutant alleles are present, but also to quantito their frequency. We previously identified somatic mosaic missense mutations in PTK3CA in three patients with Klippel-Trenaunay Syndrome (KTS). Using this molecular approach, we were able to detect recurrent mutations in KTS patients. Our results do not preclude all KTS being caused by somatic mosaic mutations in PTK3CA. We designed and validated a qPCR method for CN confirmation of their abnormal CMA results. However, there are no universal published guidelines on the design and validity of qPCR as a confirmation method in a clinical setting. A validated qPCR method for CN confirmation is emerited in aborted embryos in small samples. Expecially in clinical labs that need to be in compliance with New York State Standards of Clinical Laboratory Practice. We have designed a cost-effective, high-throughput RT-qPCR method with a quick TAT for the confirmation of CMA aberrations that are reported in our clinical laboratory. A brief description of a commercial qPCR platform addressing the challenges with 28 references run in triplicate, the proband, parents &/or sibsships (if available). We developed a multiplex assay which includes the ‘test’ gene and ‘endogenous’ gene in a combined SensiFast lo-rox master mix (Bioline) that is run in triplicate (1T). The reaction is 25µl and 1:10 serial dilution is made and ‘endogenous’ gene is run in triplicate. It is specific & specific in confirming CN for CMA with a TAT of approximately one week from design to result. By this approach, we have accurately confirmed CMA abnormalities that could not be confirmed by BAC-FISH (either because of loss of availability due to BAC size limitations). Our validation cohort consists of imbalances as small as 5kb and as large as 483kb. We propose qPCR as the standard method when confirming CMA duplications <50kb and deletions <200kb in a clinical setting.
2541T
Prenatal Testing of novel mutations of Maple syrup urine disease by next-generation sequencing. S. Chen, X. Li, H. Ge, X. Pan, F. Chen, H. Jiang, BGI-Shenzhen, Shenzhen, Guangdong, China.

Objectives: The applications of massively parallel sequencing technology to fetal cell-free DNA (cf-DNA) have brought new insight to noninvasive prenatal diagnosis. However, most previous studies based on maternal plasma sequencing have been restricted to fetal aneuploidies. We aim to report the first case to combine the target capture next-generation sequencing (NGS) for identifying new mutations in unexplained Mendelian disorders cases and NIPT of single-gene disorders by maternal plasma DNA sequencing, together to prove the feasibility and potential of clinical integration.

Methods: We recruited a pregnant couple with a maple syrup urine disease (MSUD) proband child to develop our method. Target capture next-generation sequencing was performed to identify new mutations in the related gene. The maternal plasma was isolated for DNA extraction and target sequencing with a semi custom array. Then, a haplotype-assisted strategy was developed to detect whether the fetus is inherited the pathogenic mutations in target gene. Results: We identify a heterozygosis duplication in exons 2-4 of gene BCKDHA in proband and father, a potential variant, c.392A>G in BCKDHA, in proband and mother. The parental haplotype was constructed successfully in this trio family by target region sequencing. Then, a sensitive Hidden Markov Model (HMM) was used to identify the parental transmitted allele and recombination breakpoints in the maternal plasma. In the pathogenic gene, BCKDHA, the fetal inherited the same alleles with the proband, which indicated it was also a MSUD patient. Real-time PCR and Sanger sequencing performed on DNA samples of amniotic fluid and umbilical cord blood were consistent with our diagnostic results. Conclusions: Target capture and NGS have significant efficacy and scalability for identifying new genes or new mutations in unexplained Mendelian disorders cases and NIPT by maternal plasma DNA sequencing has been proved to be feasible for noninvasive detection of single-gene disorders. This is the first report on the application of the targeted pipeline in real clinical setting, to combine the target capture next-generation sequencing for identifying new mutation and noninvasive prenatal testing by maternal plasma DNA sequencing for prenatal testing together, which indicate the potential use in routine clinical practice.

2542F

Various human SNP (single nucleotide polymorphism) markers have been found to be involved in drug metabolism. The accurate analysis of genetic variations of particular genes encoding drug metabolizing enzymes is critical for pharmacogenetics. Direct sequencing is used as a gold standard for genotyping of candidate genes, however, a faster and higher-throughput method is often required for the tests in clinical laboratories. Recently, we developed a new real-time PCR system, ExiGenotyper, with the automatic interpretation program for SNP genotyping, and genotyping kits of VKORC1 and CYP2C9, CYP2C19, and TPMT (thiopurine S-methyltransferase). SNP genotyping of VKORC1 and CYP2C9 genes is useful for determining the optimal warfarin dose. The CYP2C19 gene is a subtype of cytochrome p450 which is related with clopidogrel dosage. The TPMT gene encodes a protein that catalyses S-methylation of aromatic and heterocyclic sulfhydryl compounds, and TPMT activity can be estimated through the SNP genotyping of the TPMT gene. The allele-specific primer method was used for real-time PCR in this system for the determination of SNP genotype. For the evaluation of the system and the genotyping kits, SNP verified clinical samples by direct-sequencing were used. ExiGenotyper warfarin genotyping results of 82 human genomic DNA samples agreed with direct sequencing results. ExiGenotyper CYP2C19 genotyping results of 59 human genomic DNA samples were in concordance with direct-sequencing and ExiGenotyper CYP2C19 genotyping results of 26 samples agreed with results from a commercially available PGX device. ExiGenotyper TPMT genotyping results of 246 human genomic DNA samples were in agreement with direct-sequencing. The results suggested that the new system we developed allows accurate and rapid genotyping of SNP and is applicable in clinical laboratories.

2543W

Tay-Sachs disease (TSD) is a common autosomal recessive condition in the Ashkenazi Jewish population. It has also been observed at higher frequencies in other populations. TSD carrier screening is currently recommended by ACOG and other societies for high-risk populations, as well as for couples where at least one member is high-risk. Screening consists of both DNA analysis of the HEXA gene as well as HexA enzyme analysis. Screening has led to a significant reduction in the incidence of TSD in the high-risk populations. Today, most new cases of TSD are born to couples where at least one member of the couple does not belong to a high-risk population. According to ACOG’s position statement, ‘biochemical analysis should be used for individuals in low-risk populations’. At the time of result reporting, we noted a high percentage of indeterminate and positive TSD enzyme results without mutations present in the African American population (AA), leading us to retrospectively evaluate the data on all samples tested. Good Start Genetics’ TSD DNA sequencing test detects 87 known disease-causing mutations (including the 7.6kb deletion) plus novel truncating mutations. HexA enzyme activity was assessed on leukocytes. Out of 2656 individuals tested, 144 self-reported as AA. 16/144 (11%) were enzyme positive and 46/144 (32%) were indeterminate. In comparison, 15/1038 (1.5%) were enzyme positive and 82/1038 (7.9%) were indeterminate in self-reported Caucasians. No mutations or pseudodeficiency alleles were identified in any of the TSD enzyme positive or indeterminate AA individuals.

Based on enzyme data alone, the observed carrier frequency in our AA population is 1/9 versus an expected frequency of 1/300. Given the low incidence of TSD in AA, these data suggest that mutations could be identified in these samples, this result raises questions about the reliability of enzyme analysis in the AA population. Another clinical laboratory has also reported a higher than expected percentage of TSD enzyme indeterminate and positive results in the AA population without identifying any DNA mutations. However, only the most common mutations were tested. Our data is based on more extensive DNA analysis, and we corroborate their findings. The reference ranges for HexA enzyme activity need to be re-evaluated. Despite guidelines, enzyme activity is NOT currently an accurate method to assess TSD carrier status in the AA and possibly other populations.

2544T
Diagnostic Sequencing - implementation into routine processes. K. Stanger, T. Paprotka. GATC Biotech, Konstanz, Germany.

Next generation sequencing is on its way to play a more and more important role for analyzing genetic and genomic diseases and finding its way into clinical diagnostics. We will present the workflow of the implementation of next generation sequencing into routine processes for diagnostic and clinical purpose. Starting from the approach to design and set up pilot projects through validation studies the whole cycle till the product launch will be illuminated. Data on sensitivity and specificity from a clinical study as well as from the final product will be shown. The data sets include analysis of genetic disorders, cancer and prenatal diagnostic of more than 7,500 patients. Also, the implementation of the final test in a ISO accreditation will be described. Utilizing Next Generation Sequencing in clinical and genetic diagnosis and personalized disease risk profiling is very important in the future. Further optimization of samples isolation and preparation as well as sensitivity and data analysis pipelines will help to integrate Next Generation Sequencing into a common clinical setup.
2545F Molecular Genetic Diagnosis of Fanconi Anemia in Chinese patients. Xu Chen1, H. Liu2, W. Tang2, F. Wang3, Y. Wang4, Q. Yin5, M. Wang6, L. Guo6, P. Zhu7. 1) Department of Hematology, Peking University First Hospital, Beijing, China; 2) Molecular Medicine Lab, Hebei Yanda Hospital, Hebei, China.

Background: Fanconi anemia (FA) is a group of hereditary diseases, the typical clinical manifestations including anemia and multiple congenital malformations. However, FA patients didn’t always with malformation, thus gene mutation inspection is useful in diagnosis and differential diagnosis of FA and autoimmune aplastic anemia (AA), aplastic anemia - paroxysmal hemoglobinuria syndrome (AA-PNH). Method and Cases: FANCA, FANCC, FANC gene mutation analysis by PCR and Sanger sequencing, PHN phenotype detection by flow cytometry and chromosome breakage test was performed in 28 cases which clinically suspected FA. Results: 1) 12 cases were identified carrying FA gene mutation, 4 males and 8 females. Bone marrow failure symptoms began at 0.5 to 15 years old, with a median age of 5 years old. 2) 10 cases carrying FANCA mutations, 1 with FANCC mutations, 1 with FANC mutations. Biallelic mutations were identified in 3 cases, respectively were FANCG W184X. 3) 6 out of the 12 gene mutation-positive cases clinically with café au lait spots and/or organ malformations. Chromosome breakage test were positive in 4 out of 6 gene mutation-positive cases. 4) A large proportion of PHN cells were detected by flow cytometry in one FANCA mutated patient, with CD55 (-) 68.68% and CD59 (-) 68.81% in granulocytes. However, the patients showed multiple sporadic mutations rather than clonal mutations in PIGA gene. Conclusion: Gene mutation inspection is useful in diagnosis and differential diagnosis of FA, AA and AA-PNH.

2546W Estimating the contribution of unidentified mutations in autosomal recessive disorders. L.P. ten Kate1, M.E. Teeuw1, A. Selfani2-3, F-Z. Laaraby2, I. Hamal1, L. Henneman1, M.C. Cornel1. 1) Dept Clinical Genetics, VJ Univ Med Ctr, Amsterdam, The Netherlands, Netherlands; 2) Dept Medical Genetics, Nat. Inst. of Health, Rabat, Morocco; 3) Center of Human Genetics, Fac. of Medicine and Pharmacy, Univ. Mohammed V Souissi, Rabat, Morocco.

It has been shown before that the total pathogenic gene frequency (q) of an autosomal recessive disorder can be estimated from existing mutational data, provided that (a) the overall inbreeding coefficient (F) of patients in the sample is known and above zero; and (b) for all patients in the sample bi-allelic mutations are found (1,2). Not infrequently however, some patients reveal only one mutation or none at all even though they meet the clinical criteria. For instance, in a sample of 175 Moroccan Familial Mediterranean Fever (FMF) patients 43 were found with bi-allelic FMF mutations (27 homozygotes and 16 compound heterozygotes), 23 with only one identified mutation, and 109 without any identified mutation. Average F in this sample was 0.0122. Suppose we may assume that all 23 patients with only one identified mutation were in fact compound heterozygotes, with the compound mutation unidentified, how many of the 109 patients without an identified mutation would then be homozygotes of the same unidentified mutation? This question can be solved because there is a fixed ratio between homozygotes (H) and compound heterozygotes(C) for any given allele a, depending on F and q only, as shown in the following equation: HO/CH = [F + (1-F) a,q][2q(1-F)[1-a]], in which a represents the relative frequency of allele a, taking now the observed HO/CH ratios of identified alleles as a reference, one will be able to make a proximal estimate of the number of homozygotes with the unidentified allele (in this case 17), and subsequently the total pathogenic gene frequency including the unidentified mutation. As the HO/CH ratio will be smaller when there are two or more different identified alleles, and as some of the heterozygotes may in fact be real homozygotes, the above estimate is a maximum one. So, at least 92 of the FMF cases in this sample were not caused by recessive MEFV mutations. Other explanations for their existence are needed, (1) Ten Kate et al. J Community Genet 2010; 1: 37-42 (2) Gialluisi et al. Eur J Hum Genet 2013; Mar 13, doi: 10.1038/ejhg.2013.43 [Epub ahead of print].

2547T The effect of Long-Term Frozen Storage of Urine Samples on the detection of Chlamydia trachomatis in comparison to Preserved - Room temperature Urine Samples. M.A.K. Abdalla1, M. El-Mogry2, C. Moreira3, R. DiPietro4, I.A. Haj-Ahmad5, Y. Haj-Ahmad1,2,6. 1) Department of Biochemistry, Faculty of Science, Alexandria University, Egypt; 2) Molecular Biology Department, National Research Center, Dokki, Cairo, Egypt; 3) Brock University, 500 Glenridge Avenue, St. Catharines, ON, L2S 3A1; 4) Department of Biochemistry, Faculty of Science, McMaster University, ON, Canada; 5) Norgen Biotek Corp., 3430 Schwan Parkway, Thorton, Ontario, Canada, L2V 4Y6.

Chlamydia trachomatis is a widespread sexually transmitted disease that can be avoided by the early detection of the pathogen. Urine is a biological sample that can be considered for the detection of the pathogen. However, some studies have shown that long terms frozen urine is not suitable for molecular based assays, likely due to degradations. The main goal of this work was to compare the effect of urine chemical preservation and storage at room temperature and urine stored at 4ºC and -70ºC on the quality of the isolated DNA and hence the detection of pathogens in urine. Briefly, urine sample was spiked with known amount of Chlamydia trachomatis, and then subdivided into three sets of samples: One set of samples were stored at 4ºC, the second set at -70ºC and to the third set was chemically preserved and stored at room temperature. Total genomic DNA was isolated both at time zero and after three months. Total Urinary DNA as well as Chlamydia DNA were purified and were analyzed qualitatively using conventional and Real-time PCR. The data shows the effect of storing urine samples at different storage conditions, with or without preservative, on the quality of the isolated DNA and hence on the detection of Chlamydia from urine.


Spinal muscular atrophy (SMA) is a severe neuromuscular disorder characterized by degeneration of anterior motor neurons, resulting in a progressive muscle weakness and paralysis. SMA is caused by mutations in the survival motor neuron 1 (SMN1) gene in 5q13. The homozygous deletion of the SMN1 gene accounts for nearly 95% of SMA patients and the vast majority of parents are heterozygous carriers, as de-novo mutations occur in only 1-2% of patients. However, in a number of parents two SMN1 copies can be detected. This can either be explained by two SMN1 copies on the same chromosome 5q13 and the deletion on the other (genotype 2-0), or by a de-novo deletion and a regular SMN1 distribution in the parent (genotype 1-1). Several tests for SMN copy number analysis (e.g. MLPA, qPCR) have been developed, aiming at the identification of heterozygous carriers of the SMN1 deletion. However, they do not allow the differentiation between carriers of the 2-0 and the 1-1 genotype. Interestingly, in the patients with a de-novo deletion a significant preponderance of paternal chromosomes affected by the mutation has been observed (12 paternal de-novo deletions versus 2 maternal ones). Here we report on our experience in 118 patients with more than 200 parents of SMA patients with a homozygous deletion. All probands were investigated by MLPA; in parents with two SMN1 copies microsatellite typing was performed. In 7 parents two SMN1 copies were identified, among them four with a 2-0 genotype. Three parents had a regular 1-1 genotype (non-carriers). The de-novo deletion twice affected the maternal chromosome and once the paternal SMN1 copy. Our results and those from the literature (n=830) show that (a) about 5% of SMA parents carry two SMN1 copies and once the paternal SMN1 copy. Our results and those from the literature (n>830) show that (a) about 5% of SMA parents carry two SMN1 copies. However, they do not allow the differentiation between carriers of the 2-0 and the 1-1 genotype. In only 1-2% of patients, however, in a number of parents two SMN1 copies can be detected. This can either be explained by two SMN1 copies on the same chromosome 5q13 and the deletion on the other (genotype 2-0), or by a de-novo deletion and a regular SMN1 distribution in the parent (genotype 1-1). Several tests for SMN copy number analysis (e.g. MLPA, qPCR) have been developed, aiming at the identification of heterozygous carriers of the SMN1 deletion. However, they do not allow the differentiation between carriers of the 2-0 and the 1-1 genotype. Interestingly, in the patients with a de-novo deletion a significant preponderance of paternal chromosomes affected by the mutation has been observed (12 paternal de-novo deletions versus 2 maternal ones). Here we report on our experience in 118 patients with more than 200 parents of SMA patients with a homozygous deletion. All probands were investigated by MLPA; in parents with two SMN1 copies microsatellite typing was performed. In 7 parents two SMN1 copies were identified, among them four with a 2-0 genotype. Three parents had a regular 1-1 genotype (non-carriers). The de-novo deletion twice affected the maternal chromosome and once the paternal SMN1 copy. Our results and those from the literature (n=830) show that (a) about 5% of SMA parents carry two SMN1 copies, (b) about one third of those can be explained by de-novo deletions in the patients. We also conclude that the previously observed imbalance between paternal and maternal chromosomes affected by the de-novo deletion may be biased by the small number of cases. Finally we want to emphasize that the differentiation between the different genotypes in SMA parents (2-0 versus 1-1) does not only provide information that will assist the SMA families in further family planning but it also contributes to a more accurate risk assessment for further relatives.
A Connective Tissue Disorders NGS Panel: Development, Validation, and Novel Findings. J. Lee, M. Basehore, S. McGee, K. Kubiak, K. King, K. Champion, J. Jones, M. Friez. Greenwood Genetic Center, Greenwood, SC. Connective tissue disorders represent a heterogeneous group of more than 200 recognized conditions for which the connective tissues are the primary pathologic target. Connective tissues are the structurally supportive components of the body that form a framework or structural matrix. Connect body tissues, and provide protection of organs and storage of energy. While certain connective tissue disorders are associated with definitive phenotypic features, some patients may have non-specific or atypical presentations. Many certain connective tissue disorders are associated with definitive phenotypic features, some patients may have non-specific or atypical presentations. Many researchers have collected multiple data sets into a ‘gold standard’ for over 200 clinically heterogeneous disorders. Mutations in over 100 genes have been associated with both non-syndromic and syndromic RP. Identification of disease-causing mutations is essential for genetic counseling, carrier testing, and future reproductive planning. We have developed and validated a targeted NGS panel using traditional molecular biology techniques such as MLPA, for detecting targeted deletions or duplications, or Sanger sequencing, in cases caused by small nucleotide changes. Unfortunately, a considerable number of cases of these disorders remain unexplained, mostly because several of them are characterized by clinical and genetic heterogeneity, while in some others, as Duchenne muscular dystrophy, where around 30% are caused by point mutations, the high cost of Sanger sequencing method is unaffordable for most of the Brazilian families. In order to expand the number of conclusive molecular diagnosis, we designed a custom next-generation sequencing panel containing approximately 200 genes for MiSeq (Illumina) sequencing. These genes are associated with two main disease groups: neuromuscular disorders and skeletal dysplasias. We analyzed 26 patients: in six of them, the disease-causing mutation was known (control group), 15 had clinical symptoms compatible with neuromuscular disorders and 5 with skeletal disorders. We were able to detect the pathogenic mutations in five out of six control patients (83.3%). Among the neuromuscular patients, we identified disease-causing mutations in 11(73.3%), but we were only able to confirm the clinical hypothesis for 2 skeletal dysplasia patients (40%). The low detection rate in the later group could be explained by two main reasons: the absence of the gene associated with their clinical features in our panel and lack of sufficient coverage at the targeted region due to the limited content of our panel (e.g. TWIST2 and COL5A1). The low coverage of some exons of COL1A1 also explains why we could not detect, in the control group, all the previously known mutations. In spite of that, our experience has been positive: we identified mutations in at least 8 different genes or regions that we currently do not offer testing by Sanger sequencing, including DMD. In conclusion the results have shown that NGS is a powerful tool to improve the detection of mutations for clinical diagnosis. FAPESP/CEPID, CEGH.

Validation of clinical NGS vs. sanger sequencing: Measuring the value of orthogonal testing. S. Lincoln, S. Kash, Y. Kobayashi, G. Nilsen, J. Soranson, M. Cargill, R. Scott. InVitae, San Francisco, CA. Next Generation Sequencing (NGS) technology can help expand clinical use of constitutional genetic testing by allowing large numbers of genes to be rapidly tested in patients at low cost. However, broad adoption of NGS requires demonstration of clinical-grade accuracy, in part to address concerns about false positive findings in increasingly large gene panels. Because of these concerns many clinical NGS labs confirm pathogenic variants in patients using Sanger sequencing, adding to cost and slowing turn-around time.

We have embarked on a set of studies to measure the performance of clinical NGS by comparison with Sanger, and these studies may help determine the need for orthogonal confirmation. We apply NGS to individuals who have undergone traditional genetic testing, including both patients with various indications as well as reference DNAs from the Get-RM Program. We enrich for individuals with pathogenic variants (~50%) to measure sensitivity as well as specificity, and among those individuals we enrich for certain variation types, i.e. deleterious mutations. At the same time, one study cohort is being selected prospectively to include unbiased representation of all patients indicated for genetic testing under current clinical guidelines.

To date over 200 individuals have been tested by Illumina NGS and also by Sanger, with our NGS lab blinded to the other results. In these we see 100%; concordance for both variant and reference bases called. We plan to expand these studies to include about 1000 samples by the time of the ASHG 2013 meeting and 2000 thereafter.

We are also augmenting the study with 9 HapMap samples, for which we have collected multiple data sets into a ‘gold standard’ for over 200 clinically relevant genes. These samples allow many variant and invariant positions to be studied per sample, with the caveat that known pathogenic variants are found in a frequency lower than 1%. NGS data quality depends on careful choice of platform, assay, algorithms, parameters and QC criteria. We will describe our methods in addition to the validation study details. We intend to quickly make as much of the above data public as possible.

Clinical utility of next generation sequencing for the molecular diagnosis of genetically heterogeneous retinitis pigmentosa. J. Wang1, V.W. Zhang1, F.Y. Li1, C. Truong1, G. Wang1, P.W. Chiang5, R.A. Lewis1*, L.J. Wong1. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Casey Eye Institute, Oregon Health and Science University, Portland, OR; 3) Department of Ophthalmology, Baylor College of Medicine, Houston, TX.

Background: Retinitis pigmentosa (RP) is one of the most genetically heterogeneous disorders. Mutations in over 100 genes have been associated with both non-syndromic and syndromic RP. Identification of disease-causing mutations is essential for genetic counseling, carrier testing, and for future reproductive planning. RP-causing genes are chronologically divided into multiple genes for RP by a high throughput next generation sequencing (NGS) is a cost-effective approach to that goal. Methods: We developed a target gene capture sequencing (TCS) approach with the NimbleGen solution-based capture design of 66 genes currently known to cause RP, followed by NGS analysis on Illumina HiSeq2000. Results: A total of 202,800 bp of target sequences including 939 coding exons and 20 bp of flanking intron regions were sequenced to an average depth of 700X per base. Six consistently pathogenic covered exons were completed by PCR/Sanger sequencing to ensure 100% coverage for the entire coding regions. The phase I validation sample has a total 135 variants being identified, which are 100% concordant with Sanger sequencing results. Phase II validation was performed on 12 samples and known mutations in 4 different genes: ABCA4, RP1, RP2, and USH2A. All previously detected disease-causing mutations were identified correctly. We also have analyzed 25 samples from unrelated individuals with pigmentary retinal dystrophies, LCA, or RP-related disorders. Deleterious mutations were detected in 19 patients in 19 genes due to autosomal recessive, dominant and X-linked inheritance. Our positive detection rate is 76%.

In addition to single nucleotide substitution and small indels, homozygous insertion of a 355 bp Alu sequence in exon 10 of the MAK gene in two unrelated patients, and a homozygous exonic deletion in the EYS gene in one patient were detected in this cohort. Conclusion: Our data underscore the importance clinical utility of NGS-based analysis in the molecular diagnosis of RP. A strategic data analysis has proven accurate for mutation- and variant-identification and annotation. The deep exonic base-to-base coverage of all coding regions plus Sanger confirmation allow the accurate identification of all point mutations, small indels, and even large Alu insertions, which is difficult to detect from NGS data. The TCS approach greatly improves the diagnosis of RP in a cost and time efficient manner.


The Human Genome Research Center (HGRC) offers molecular diagnosis to several diseases, mainly neuromuscular disorders and some cases of skeletal dysplasias using traditional molecular biology techniques such as MLPA, for detecting targeted deletions or duplications, or Sanger sequencing, in cases caused by small nucleotide changes. Unfortunately, a considerable number of cases of these disorders remain unexplained, mostly because several of them are characterized by clinical and genetic heterogeneity, while in some others, as Duchenne muscular dystrophy, where around 30% are caused by point mutations, the high cost of Sanger sequencing method is unaffordable for most of the Brazilian families. In order to expand the number of conclusive molecular diagnosis, we designed a custom next-generation sequencing panel containing approximately 200 genes for MiSeq (Illumina) sequencing. These genes are associated with two main disease groups: neuromuscular disorders and skeletal dysplasias. We analyzed 26 patients: in six of them, the disease-causing mutation was known (control group), 15 had clinical symptoms compatible with neuromuscular disorders and 5 with skeletal disorders. We were able to detect the pathogenic mutations in five out of six control patients (83.3%). Among the neuromuscular patients, we identified disease-causing mutations in 11(73.3%), but we were only able to confirm the clinical hypothesis for 2 skeletal dysplasia patients (40%). The low detection rate in the later group could be explained by two main reasons: the absence of the gene associated with their clinical features in our panel and lack of sufficient coverage at the targeted region due to the limited content of our panel (e.g. TWIST2 and COL5A1). The low coverage of some exons of COL1A1 also explains why we could not detect, in the control group, all the previously known mutations. In spite of that, our experience has been positive: we identified mutations in at least 8 different genes or regions that we currently do not offer testing by Sanger sequencing, including DMD. In conclusion the results have shown that NGS is a powerful tool to improve the detection of mutations for clinical diagnosis. FAPESP/CEPID, CEGH.
2553T

Evaluation of fragile X screening methods for early detection of affected infants. P. Mueller, J. Lyons, G. Kerr. Newborn Screening and Molecular Biology Branch, CDC, Atlanta, GA.

Purpose: Expanded CGG repeats (>200 repeats) in the FMR1 promoter are known to cause Fragile X in males; premutations (55 to 199 repeats) cause fragile X-associated primary ovarian insufficiency in females and tremor/ataxia syndrome in both males and females. Since the diagnosis in these expanded repeats is near 100%, the larger CGG repeats are very difficult to PCR amplify while normal repeats amplify preferentially. Therefore, it is difficult to distinguish between normal homozygous females (20% to 40% of females), and those with one normal allele and a CGG expansion. The goal of this study was to evaluate and modify, as needed, published methods for detecting CGG repeat expansions in the FMR1 gene promoter region for potential use as a Fragile X screening assay. Methods: We have evaluated and tested modifications of PCR-based screening methods including those using CGG targeted primers, meltpoint curve, and heat pulses during PCR extension to facilitate the detection of large CGG repeats. Results: Our analysis of 80 normal, 3 gray-zone, and 38 pre- and full mutation samples using the protocol by Tassone that includes a CGG targeted primer (J Mol Diagn, 10:43-49, 2008) called all correctly except one female heterozygous normal and expanded sample. The Orpana (Anal Chem, 84:2081-2087, 2012) heat pulse method using agarose gel electrophoresis detects pre- and full mutation males since there is no visible product for the normal allele even if the amplicon from an expansion is not evident. Female samples with a normal expansion are less evident. The Teo method (Clin Chem 58:568-570, 2012) using a melting curve analysis of triplet-primed PCRs in both the 5’ and 3’ directions resulted in optimal temperature cutoffs and the Teo assay for single nucleotide variants with an additional optimization in each laboratory. The 3’ assay gave the more robust optimal combination of sensitivity and specificity for pre- and full mutations. Conclusions: The Tassone method gave the best performance for both males and females, while the Orpana method gave the best result for females. The Teo method was more robust than the 5’ assay. A cost analysis based on reagent and supply cost with amortization of major equipment over 5 years, assuming 55,602 samples/year (median of annual state births, 2010), indicated that the Orpana method was different from the published method necessitating optimization in each laboratory. The 3’ assay proved the more robust optimal combination of sensitivity and specificity for pre- and full mutations. Optimized conditions are now available.

2555W


Next-generation sequencing (NGS) has the great potential to improve clinical care for children to diagnose rare Mendelian diseases as well as for adults and provide specific treatments such as for various cancers. Based upon medical need of individuals receiving clinical care at Montefiore Medical Center, we designed a custom gene panel, spanning 5 Mb and consisting of 650 genes targeted known Mendelian loci, some pediatric diseases and some loci with mutations in various cancer types. To determine the sensitivity and specificity of our custom panel we also used several prespecified panels (Ion Torrent and Whole Exome Sequencing (WES) on cancer samples. The gene panel was created using the Roche-NimbleGen SeqCap V3 capture system. We sequenced over 300 Rapid Run technology. A total of 24 DNA samples were analyzed, including 7 control DNA samples with known variants. Sequence data was annotated and variants were identified through the Wiki-based Automated Sequence Processor (WASP) analytical pipeline (PMID: 22942039) and customized analyses using the Broad Institute Genome Analysis Tool Kit. Of the 17 subjects (15 unrelated, 2 related), eight had known Mendelian disorders, eight had pediatric diseases and one had breast cancer (tumor/normal). We identified likely causative variants in two unrelated subjects that included mutations in CASQ2, p.Gln67*, c.532+1G→A (Ventricular Tachycardia, Catecholaminergic Polymorphic, 2; OMIM#611938) and TP53, p. Arg213* (Breast Cancer; OMIM#114480), two related subjects (siblings) that included mutations in LDB3, p.V118M (Cardiomyopathy, Dilated, 1C; OMIM#601493) and MAP3K9, c.3421-1G→A (Malignant Hyperthermia, 1B, OMIM#600598). We identified two breast cancer samples, two additional targeted cancer panels were used (AmpliSeq Cancer Hotspot and the Comprehensive Cancer Panel using the Ion Torrent Technology). In addition, we performed WES on the breast cancer sample and could not detect the known hotspot mutations, identifying a novel nonsense variant in the TP53 tumor suppressor gene. This variant was detected on all platforms. We found targeted capture high-throughput sequencing to be a cost effective, time sensitive and an efficient approach in detecting pathogenic and likely causative variants to aid the diagnosis of individuals in our Bronx patient population.

2554F


Members of the Ashkenazi are at elevated risk for a number of inherited genetic syndromes. Widespread and successful adoption of carrier screening among the Ashkenazi Jewish community has greatly reduced the incidence of these inherited diseases. We have developed an expanded screening panel for 51 mutations across 18 diseases prevalent in this community using the Illumina, Inc. BeadXpress Platform. This panel comprises 8 multiplex PCR amplification reactions and subsequent allele specific primer extensions pooled for the detection of 103 independent wild type or mutant signals at once via VeraCode™ technology. Due to widespread adoption of carrier screening in the Ashkenazi community and population dispersion, a significant percentage of recently born disease affected individuals have either uncommon or de novo mutations. As such, we have included the addition of several less common variants across multiple diseases and will track incidence of these for justification of additional rare variants of pathological importance in future panels. This laboratory developed test was will track incidence of these for justification of additional rare variants of pathological importance in future panels. This laboratory developed test was

2556T


Clinical whole exome sequencing (WES) coupled with mitochondrial genome sequencing (MGS) was performed on 100 cancer WES cases. Clinical reports have been issued from over 1000 approximately 1700 cases including about 1600 germline WES cases and 800 cancer WES cases. We have received clinical WES cases from over 300 cancer patients and 800 germline WES cases. Our WES and MGS projects include approximately 1700 cases.

Clinical reports have been issued from over 1000 approximately 1700 cases including about 1600 germline WES cases and 800 cancer WES cases. We have received clinical WES cases from over 300 cancer patients and 800 germline WES cases. Our WES and MGS projects include approximately 1700 cases.

Clinical reports have been issued from over 1000 approximately 1700 cases including about 1600 germline WES cases and 800 cancer WES cases. We have received clinical WES cases from over 300 cancer patients and 800 germline WES cases. Our WES and MGS projects include approximately 1700 cases.
Introduction: Germline mutations in BRCA1 and BRCA2 are known to increase the risk for developing familial breast and ovarian cancer. The large genomic rearrangements (LGRs) involving the BRCA genes have been reported to account for variable proportion of familial breast cancer patients according to ethnic populations. Here, we studied the contribution of LGRs in BRCA1 and BRCA2 to high-risk breast cancer patients in Korea. Methods: A total of 243 patients with two or more familial history of breast cancer (BC) or one or more familial history of ovarian cancer (OC) were enrolled in the study. All patients were screened by direct sequencing (n=177), or mutation scanning like fluorescence-based conformation-sensitive gel electrophoresis (F-CSGE) or denaturing high performance liquid chromatography (DHPLC) and/or direct sequencing (n=66). Multiplex ligation-dependent probe amplification (MLPA) was done in 144 patients who were mutation negative for the BRCA genes. Results: BRCA1 mutations were detected in 48 patients and BRCA2 mutations in 53 patients. One patient was harboring both BRCA1 and BRCA2 mutations. Among 144 BRCA mutation negative patients, BRCA1 LGRs were identified in three one and no BRCA2 LGR was identified. In addition, the BRCA1 LGRs account for 6.25% (3/48) of all BRCA1 mutations in this population. Conclusion: This study suggests that MLPA might be considered in BRCA1/2 mutation negative familial BC or OC patients in Korea, although contribution of LGRs is low in this population.


Bilateral sensorineural hearing loss (BSLNHL) has significant genetic heterogeneity with more than 200 associated genes. Current testing for BLSNLH is either done by Sanger sequencing or next generation sequencing (NGS) of gene panels. These tests can be cumbersome and are not comprehensive. We report the results of applying exome sequencing for the diagnosis of BLSNLH. This work was carried out through the CHOP/UPenn Pediatric Genetic Sequencing (PediSeq) project, an NHGRI funded grant, to study the application of exome and genome sequencing to clinical care. Samples from 30 affected individuals with BLSNLH (10 positive controls and 20 unknown samples) were analyzed following exome capture (Agilent SureSelect ton) and two duplications (448 kb) that overlay an SHOX genes associated with idiopathic and syndromic short stature, genes suggestive of a potential role in stature, and the PAR1 area including SHOX. An IRB-approved retrospective effort to identify ISS cases involved chart review of demographics, clinical presentation, lab data and imaging results from patients referred to Mayo Clinic Rochester for short stature between 2009 and 2012. We classified 200 patients with short stature. A total of 22 (11%) constitutional delay of growth and puberty, 23 (11%) nutritional, 27 (13%) identified genetic syndrome, 19 (9%) syndromic short stature with no identifiable cause, 12 (6%) endocrinopathy, 12 (6%) chronic systemic disease, 8 (2%) intrauterine growth retardation and 14 (7%) short stature due to miscellaneous causes. The ISS patients were approached for consent to perform custom chromosomal microarray testing. A prospective effort also assessed ISS patients from pediatric endocrinology and medical genetics at Mayo Clinic Rochester. Finally, cases referred for short stature that were normal by clinical testing using a chromosomal microarray not specifically targeting stature genes were also included. To date, these combined recruitment methods have resulted in analysis of 224 cases. The identification of PAR1 deletions (1.3 Mb and 1.9 Mb) encompassing SHOX and two duplications (448 kb) that overlap an evolutionarily conserved region near SHOX of possible significance. This study is ongoing with a recruitment target of 200 patients with ISS.

Tay-Sachs disease (TSD) is an autosomal recessive, neurodegenerative disorder caused by mutations in the HEXA gene, which lead to deficient activity of the Hex A enzyme and the accumulation of its substrate, ganglioside in neural cells. Hex A is comprised of two subunits, α and β encoded by HEXA and HEXB respectively. A related isoyme, Hex B consists of two β subunits. To prevent TSD in high-risk populations, measurement of Hex A enzymatic activity has been routinely used as the primary method of carrier screening. This assay utilizes a synthetic substrate, 4-MUG, which can be hydrolyzed by both Hex A and Hex B, in combination with a heat-inactivation step, to differentiate the activities of the heat-labile Hex A from heat-stable Hex B in fluids or blood cells. TSD carriers, noncarriers, and affected individuals can be identified by their different ranges of Hex A%. Here we describe four TSD carriers identified by mutation analysis whose Hex A% results from enzymatic activity assay in leukocytes showed that one was a noncarrier (Hex A%: 68.4%) and the other three were within the inconclusive range (avg. Hex A%: 53.0±2.7). HEXB gene sequencing revealed a c.[1627G>A; “81.82delTG] genotype in the first case (noncarrier by Hex A%) and a c.[619A>G(+)*81_*82delTG] genotype in the other three (inconclusive by Hex A%). An additional 41 TSD carriers confirmed by both mutation analysis and Hex A% (avg. Hex A%: 43.1±4.7) were screened for these two genotypes. None of them had c.[1627G>A; “81.82delTG] genotype and the allele frequency of c.[81A>G(+)*81_*82delTG] genotype in these carriers was 9.8% (8/82), different from the general population. These results, therefore, suggest that these two genotypes are strongly associated with atypical Hex A% findings as all four mutation positive individuals who also had these polymorphisms did not have enzyme results in the carrier range. The 1627G>A and the *81_*82delTG polymorphisms have been previously found in TSD carriers whose Hex A% results from enzymatic activity assay in leukocytes showed that one was a noncarrier (Hex A%: 68.4) and the other three were within the inconclusive range (avg. Hex A%: 53.0±2.7).

Marfan syndrome (MFS, MIM#154700) is a connective tissue disorder with variable clinical manifestations in the skeletal, ocular and cardiovascular systems. Aortic dilation, dissection and mitral valve prolapse are common in MFS patients. However, aortic root dilation may not be noted without aortic insufficiency or rupture, as one of leading causes of premature death in MFS patients. Clinical WES performed in the Medical Genetics Laboratories at Baylor College of Medicine identified a heterozygous c.3590G>A (p.R1170H) mutation in the FBNT1 gene in a 2 year old patient without known cardiovascular issues, who was diagnosed with cardiovascular-renal hypoplasia. To our knowledge, this is the first patient with the p.R1170H mutation in FBNT1 exhibiting aortic root dilation at a young age. Based on above results, close monitoring is needed for cardiovascular complications and intervention.

In summary, the case reported here demonstrates the benefit and challenges of returning incidental findings with proper consent, thorough communication and deliberate follow-up evaluation. Collective information and additional studies could broaden the phenotypic spectrum of the p.R1170H mutation in the FBNT1 gene.
2564W
Establishment of Molecular Diagnostic Platform for Leber Congenital Amaurosis using Extensive Multi-Gene Panel Sequencing. S.H. Seo1, Y.S. Yu2, J.M. Hwang2, H. Park1, S.I. Cho3, S.S. Park1, M.W. Seong1. 1) Departments of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea; 2) Departments of Ophthalmology, Seoul National University Hospital, Seoul, Korea; 3) Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, Korea.

Introduction: Leber congenital amaurosis (LCA) leading to blindness is genetically heterogeneous, and thus confirming the molecular diagnosis can be challenging. So far, more than 10 genes have been discovered to be associated with this disease, but these genes account for a small proportion of LCA in this country. Herein, we developed an extensive diagnostic multi-gene panel including more than 200 retinal dystrophy-associated genes.

Methods: Twenty LCA patients were participated in the study, including 2 patients with previously defined pathogenic mutations, in order to verify the performance of our diagnostic platform. A total of 203 genes were analyzed by enrichment of all coding regions and flanking intronic regions with a customized NimbleGen SeqCap EZ Choice library (Roche) and sequencing with a HiSeq (Illumina). All pathogenic variants discovered and low-coverage regions under 10X of coverage depth were resequenced by Sanger method.

Results: On average, the coverage depth was 800X per base pair (bp), and only less than 1% of the regions showed coverage under 10X. Two patients for assay validation were successfully confirmed using multi-gene panel sequencing. Among remaining 18 patients, two pathogenic mutations were detected in 5 patients including two novel LCA genes as well as known LCA genes like CEP290 and RPGRIP1. In addition, 8 patients, only single possible pathogenic mutations were found. Conclusion: Our extensive diagnostic platform has identified novel genetic causes in LCA patients. We demonstrated the successful use of massive parallel sequencing in confirming the molecular diagnosis of LCA patients, and that multi-gene panels can be used with more advantages over exome sequencing in the clinical fields.

2565T
Detection of low-level mosaic microdeletion in Neurofibromatosis type 1
J. Xie, A. Poplawski, C. Fu, T. Callens, J. Williams, H. Zhan, L. Messiain. Dept of Genetics, Medical Genomics Laboratory, University of Alabama at Birmingham, Birmingham, AL.

Neurofibromatosis type 1 (NF1; MIM#162200), an autosomal dominant neurocutaneous disorder affecting ~1 in 3000 individuals worldwide, is caused by mutations in NF1 gene. The NF1 gene has a high mutational rate compared to other disease-related genomic loci. As many as 30-50% of NF1 patients present as sporadic cases, i.e. do not have a parent affected by the disorder. A fraction of these ‘founder’ cases presents with mosaicism for a NF1 mutation, as a consequence of de novo mutations arising postzygotically during embryonic development. Mosaicism is an important consideration in NF1, given its impact on the clinical phenotype and transmission risks. Estimation of the frequency of mosaicism in sporadic NF1 is however challenging due to the notorious variability of the phenotype, the complexity of the NF1 mutational spectrum and the limitations of the current techniques to accurately detect low-level mosaicism for all of the potential mutational targets in the NF1 gene. A first estimation on the frequency of NF1 mosaicism was provided through the study of patients carrying an NF1 microdeletion as identified by a transferase dependent Probe Amplification (MLPA) and confirmed using FISH (Messiaean et al. Hum Mutat. 32, 231, 2011). The frequency of 10% in this cohort necessarily is an underestimate, as low-level mosaic microdeletions would have been missed in the initial assessment. We recently identified a mosaic NF1 patient carrying an NF1 microdeletion in only 8% blood cells (4% of the alleles). Such low-level mosaicism will escape detection using MLPA or aCGH. Therefore, in order to facilitate more sensitive and accurate molecular diagnosis for NF1 microdeletions, we developed ddPCR. ddPCR allows measurements of smaller fold changes compared to MLPA or aCGH. TaqMan® copy number assay probes were used for the target NF1 and reference RNase P genes and absolute measurements were obtained using the droplet reader. Our data showed that ddPCR can detect mosaic NF1 microdeletion in as low as 5-10% blood cells (p<0.05). As upon transmission, the NF1 microdeletion tends to result in a more severe phenotype with earlier age of onset in the constitutionally affected patient. Sensitive detection of this specific type of mutation in ‘founder’ patients is warranted as a part of a comprehensive assessment.

2566F
Incorporating new disease genes into clinical whole exome sequencing (WES): annotation update, interpretation challenges and customized laboratory medicine. Yang1, F. Yao1, Z. Yu1, Y. Liu1, J.Z. Niu2, M.S. Leduc1, R.E. Person1, M.T. Hardison1, J. Zhang1, M. Bainbridge3, J.G. Reid2, A.C. Hawes2, Y. Ding2, A.A. Braxton1, P.A. Ward4, M.L. Landsverk1, A. Willis5, D.M. Muzny1, S.E. Pion1, J.R. Lupsik1, L. Beaudet1, R.A. Gibbs1, C.M. Eng1. 1) Baylor Whole Genome Laboratory (WGL), 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX.

The disease genes list is accelerated rapidly with the development of NGS technologies, availability of large-scale biology data and accumulated knowledge of gene function. Hundreds of new disease genes are identified each year. Translating new research discoveries into clinical laboratory services is a challenging task as accurate and timely manner can lead to improved diagnostic sensitivity and better understanding of genotype-phenotype correlations. For example, of the 265 positive cases diagnosed by clinical WES at Baylor WGL during the period of January 2012 through May 2013, 13% harbor causative mutations in disease genes discovered since January 2012. WES testing is an ideal platform for timely incorporation of new disease genes since it interrogates the entire coding regions of the genome, making it possible for updating disease gene annotation at any time during or even after the analysis process. However, extensive literature review to verify new Mendelian disorders and disease-causing mechanisms must be performed.

The clinical laboratory should also make every effort to facilitate familial mutation studies for new disease genes, which usually do not yet have single-gene clinical tests available (e.g., Sanger sequencing). In this case, we reported a patient with mutations in MEGF8 (gene published November 2012, WES diagnosis reported the same month), another patient with a mutation in HADC8 (case signed out July 2012 as negative, gene published May 2012, diagnosis updated the same month), and recurrent diagnoses from disease genes (MAGEL2, ASXL3) discovered by our laboratory. Follow up communications with referring physicians are conducted and customized targeted mutation analyses, including prenatal testing when applicable, are made available on a clinical basis for family members.

2567W
Implementation of a Quality Assurance Program for Next Generation Sequencing (NGS) Testing
J. J. Eng1, C.A. Stolle1, L.C. Davidson1, M.J. Italia2, J.R. Lupski1, Z. Fan1, T.Tischler1,1) Department of Molecular and Human Genetics, University of Texas Health Science Center at Houston, Houston, TX; 2) Center for Medical Informatics, Children’s Hospital of Philadelphia, Philadelphia, PA.

Clinical laboratories are governed by regulations from Clinical Laboratory Improvement Amendments (CLIA) and many also follow guidelines and multi-gene panels from the College of American Pathologists (CAP). These regulations are in place to ensure high standards and reliable test results in the clinical environment. Next generation sequencing (NGS) has revolutionized genomics and is now being increasingly applied to clinical molecular diagnosis. Despite this increasing popularity, implementation of NGS in a clinical environment is challenging due to the inherent complexities of the rapidly evolving NGS technologies and the accompanying requirements for sophisticated data analysis. Moreover until recently, few guidelines or quality assurance programs have been put in place to meet regulatory standards. A workflow management quality control system (Pegasus) was used to support a customized data management, processing, and analysis workflow. In summary, we have developed a comprehensive quality management program in the laboratory for NGS based tests. This program, which has been implemented in our clinical practice, will help meet regulatory standards as well as assure the integrity and quality of NGS results.

Cell free DNA (cfDNA) offers a non-invasive diagnostic approach to a wide range of clinical disorders. Next-generation sequencing (NGS) has enabled the detection of unknown fetal genomic variations from cfDNA of maternal plasma. However, due to the lengthy sequencing time, some platforms are still not suitable for clinical application. The relatively low concentration of cfDNA in the blood has presented many challenges as well. By using targeted sequencing, only a subset of genes or a defined region in the genome is sequenced, which allows high sequence coverage critical for identifying rare genetic variations. The purpose of our study was twofold, first, to sequence and analyze the genomic coverage of cfDNA using NGS technologies and second, to test the sensitivity and accuracy of detecting the genomic mutation from cfDNA by using Ion AmpliSeqTM Inherited Disease Panel (IDP), which enables multiplexed target selection of exons of 328 genes implicated in genetic disorders. Whole blood (6 mL) was collected from healthy donors, and plasma was separated by centrifugation. The cfDNA and genomic DNA (gDNA) were extracted using column-based methods. The concentration of cfDNA was 8.9 ± 0.5 ng/µl (30 ng/ml of blood), which was in agreement with reported values in healthy controls (5-30 ng/ml). Libraries were prepared accordingly: 1) Ion Xpress TM Plus cDNA Fragment library kit was used for sequencing with the Ion Torrent PGM™ platform with 316 chips; 2) ThruPLEXTM FD Prep Kit (Rubicon Genomics) was used for Illumina® NGS platform. Result showed that the genomic coverage of cfDNA sequencing was 97% of that of the genomic DNA, given the same number of reads. The results were similar in both Illumina® and Ion Torrent PGM™ platforms. The study also showed that the cfDNA present in the blood stream is representative of the global content of the gDNA. To study the feasibility of detecting the genomic variation of cfDNA, semiconductor sequencing will be performed by using IDP. To mimic the condition of maternal blood, cfDNA samples will be mixed in the range of 5 to 20%, which is the ratio of cell free fetal DNA in the maternal blood. Through this study, we will determine the possibility of using cfDNA from blood for detection of genomic mutations. This suggests that the use of cfDNA in conjunction with accurate and rapid semiconductor sequencing technologies could be the next step forward in developing non-invasive diagnostic testing.

2569F A systematic approach to assessing the clinical significance of genetic variants. H. Duzkale1,2, J. Shen1,2, H. McLaughlin1,2, A. Affares1,2, M.A Kelly2, T.J Pugh1,2, B.H Funke1,2, H.L Rehm1,2, M.S Lebo1,2. 1) Harvard Medical School, Boston, MA; 2) Partners HealthCare Center for Personalized Genetic Medicine, Cambridge, MA.

Molecular genetic testing can improve the accuracy of diagnosis, prognosis, and risk assessment for patients and their family members. Recent advances in low-cost, high-throughput DNA sequencing technologies have enabled the rapid expansion of genetic tests, as well as an expansion of variant- and gene-level data and their associations with human disease. The number of variants assessed in our laboratory has tripled in the past two years through launching next generation sequencing (NGS) gene panels. While NGS panels have increased the analytical and clinical sensitivity of our assays, the increased content has also added a significant burden of interpretation and higher rate of variants of uncertain significance (VUS). Our laboratory has developed a systematic approach to accurately, efficient and timely assessment of variants for pathogenicity over the past decade. Using our semi-automated tool, the average time for a clinical variant assessment has been reduced to 22 min. We first gather and validate variant information from publications and clinical notes, from internal, collaborative and public variant databases, and from various bioinformatics resources; we then perform statistical analyses, evaluation of research and functional data, and computational predictions on the data to determine the likelihood of pathogenicity; finally, we weigh all the evidence to reach an overall interpretation on the potential for each variant to be disease-causing. In this report, we highlight the principles of variant assessment, address the caveats and pitfalls, and provide customizable tools and examples to illustrate the process. We also demonstrate the importance of variant reassessment, as we recently downgraded 11 of 106 (10%) pathogenic or likely pathogenic cardiomyopathy variants to VUS, mostly due to their presence in large population studies. Our laboratory has evaluated 245 genes in 53 diseases in a total of 22,490 cases since its inception. 17,245 variants have approved clinical classifications and 7,555 were validated and reported in patients. We have cataloged and submitted 7,129 variants to ClinVar database, an initiative to publicly share genetic test results as well as phenotype information of patients. Of these variants, 40% have associated rsIDs in dbSNP. 20% have been reported in the literature, and 40% were novel. By sharing our experience, we hope to bring the clinical and research communities together to build a framework for variant assessment.

2570W Assessment Of Detection Of Proviral DNA and RT (MET 184 VAL) Gene Resistance Mutation in HIV-1 Identified by Multiplex PCR and Restriction Fragment Digestion Assay. R. SHRESTHA1,2, S. KHADKA1, S.R. WAGLE3, A. SAPKOTA4. 1) Center for Molecular Dynamics Nepal (CMDN), Intrepid Nepal, Pvt. Ltd. Thapathali-11, Kathmandu, Nepal; 2) Department of Laboratory Medicine, Nobel Hospital and research Institute, Kathmandu, Nepal.

ABSTRACT Proviral DNA and ART resistance forms the corner stone of a short chemotherapy course for Post exposure prophylaxis. Potent antiretroviral therapies that suppress cell-free plasma viral RNA levels below the limit of current assay detection necessitate other complementary approaches for assessing viral burden, such as quantification of cell associated proviral DNA. The quantification of human immunodeficiency virus type-1 (HIV 1) RNA is the cornerstone for monitoring the effectiveness of antiretroviral therapy. For the majority of therapy naive patients, three to six months of potent antiretroviral therapy usually suppresses cell-free HIV 1 RNA levels to less than 50 copies/mL of plasma. Objective: To rapidly identify the proviral DNA and To study mutation in specific part (meth184val; ATG - GTA substitution) of M184V gene of HAART resistance in HIV-1 Reverse Transcriptase. Method: An analytical study was design to explore the role of proviral DNA and RT (M184V) gene mutation identified by PCR and RFLP. Results and Discussion: The male patients outnumbered female. Among 13 male subject 2(15.4%) were proviral DNA positive and 11(84.6%) were proviral negative. Among 2 female subject 1(50.0%) were proviral DNA positive and 1(50.0%) were proviral negative. All isolates were only digested by CVIAll restriction enzyme and confirmed as wild type. No mutation was detected in the analyzed sample. Conclusion: In summary, we demonstrated here that the RT gene (specific part) proviral DNA load was below the limit of detection in patients. Since, we can say that the incidence of developing of resistance virus is in increasing order. In addition, Potent antiretroviral therapies suppress cell-free plasma viral RNA levels below the limit of current assay detection necessitate other complementary approaches for assessing viral burden, such as quantification of cell associated proviral DNA. Therefore, it is recommended to use this technique for the diagnosis of Drug Resistance Mutation in HIV-1 and the range of ART for their Better Health Care. Keywords: HIV/AIDS, HAART, Drug resistance, mutation, env gene, RT gene.
Posters: Clinical Genetic Testing

2571T
International external quality assessment for diagnostic Next Generation Sequencing. S. Absa1, J. Coxhead2, P. Westwood3, K. Thomson4, H. Scheffer5, S. Bhaskar6, G. Taylor7, Z. Deans8, S. Patton9, 1) Genetics Laboratories, Cambridge University Hospitals, Cambridge, United Kingdom; 2) Newcastle University, Newcastle, UK; 3) Western General Hospital, Edinburgh, UK; 4) Oxford University Hospital, Oxford, UK; 5) University Medical Center Nijmegen, Nijmegen, the Netherlands; 6) Central Manchester Foundation Trust, Manchester, UK; 7) University of Melbourne, Victoria, Australia; 8) UK National External Quality Assessment Scheme for Molecular Genetics, Edinburgh, UK; 9) European Molecular Genetics Quality Network, Manchester, UK.

Next Generation Sequencing (NGS) is increasingly being introduced into clinical genetics laboratories worldwide. The huge amount of data generated by NGS cannot be duplicated by alternative methods for laboratories to internally validate all results, therefore external assessment of data is required. The UK National External Quality Assessment Scheme (UKNEQAS) for Molecular Genetics and the European Molecular Genetics Quality Network (EMQN) have developed a joint EQA scheme for NGS, with the aims to: (a) assess and improve quality; (b) enable laboratories to benchmark their NGS service against others and against best practice; (c) work towards consistency of reporting clinical results generated by NGS; and (d) contribute towards best practice. EMQN and UKNEQAS offer numerous disease specific EQA schemes, and the challenge for developing NGS EQA was to ensure it does not duplicate what is already available, making it generic (independent of genes, diseases, and platforms) and applicable to as many users as possible. A survey of 1020 labs worldwide in Dec 2011 generated replies from 52% about methods, platforms, and loci being tested. In the diagnostic experience of NGS varied from less than a year (76% labs) to over 3 years (1.9%). All respondents expressed an interest and need for NGS EQA. The survey results were used to develop a follow-up survey to a subset of labs, and then establish a pilot EQA. 30 labs were selected on the basis of diagnostic experience with NGS and covering a variety of technology platforms. These labs were sent a genomic DNA sample and asked to sequence either the smallest gene panel or largest single gene which the lab tested, submit technical details, and genotypes at known SNPs. The DNA was validated in 3 diagnostic labs and by 3 NGS platform manufacturers. The number of genes sequenced per lab varied from 1-625. 1011 variants were reported against reference sequences, in a total of 145 genes. 30 genes were sequenced by more than one lab, allowing comparison of results between participating laboratories. The number of genes sequenced per lab varied from 16x to 100x. Disease specific EQA has drastically improved the quality of results and consistency in diagnostic reports. This NGS EQA will play an important role in enabling labs to benchmark this new technology, assess the accuracy of data and facilitate high quality reporting for patient benefit.

2572F
Development of a rapid and comprehensive genetic testing service for nephrotic syndrome using next generation sequencing. E.J. Ashton1, D. Bockenhauer2, N.J. Lynch1, 1) NE Thames Regional Genetics Service, Great Ormond Street Hospital, London,WC1N 3BH United Kingdom; 2) Department of Nephrology, Great Ormond Street Hospital for Children, London, WC1N 3JH.

Nephrotic syndrome is a renal disorder presenting with proteinuria, oedema and hyperalbuminemia. Mutations in at least 18 different genes are known to be responsible for nephrotic syndrome and genetic testing has been a long and time consuming process as each gene would be sequenced independently until a mutation could be identified. Identification of a genetic basis provides an explanation for the disease, enables genetic and prognostic counselling and increasingly affects management of the disease. The advent of next generation sequencing has meant that it is possible to screen a large number of genes simultaneously, which would previously have been impractical and unaffordable. We have developed a targeted re-sequencing screen comparing two different library preparation methods (the Haloplex™ target enrichment system from Agilent and TruSeq Custom Amplicon from Illumina) and used these to screen for mutations in up to 18 genes known to be involved in nephrotic syndrome. Sequencing was carried out using the Illumina MiSeq system using both 2x150 bp and 2x250 bp sequencing chemistries and data analysis was performed using NextGENEd software from SoftGenetics. All mutations detected by Next Generation Sequencing were confirmed using Sanger sequencing. We have screened over 60 patients so far who were referred with a clinical diagnosis of steroid-resistant nephrotic syndrome aged from newborn to adulthood, using a combination of the two different library preparation methods, with the majority of mutations so far being identified in the NPHS1 gene. In cases where we were unable to identify a definite pathogenic mutation (or mutations) we considered that this may be due to mutations in regions of a gene not covered by our design, an uncertain clinical diagnosis or mutations in genes yet to be identified in nephrotic syndrome and therefore not included in our designs. The flexibility of the library designs means that it is simple to add in genes newly associated with nephrotic syndrome. We have developed a rapid and comprehensive genetic testing service for nephrotic syndrome, which is being used to screen patients with steroid-resistant nephrotic syndrome for at least 18 genes of interest. It is anticipated that this simple and cost effective targeted re-sequencing method will lead to improved detection rates combined with a much reduced cost and turnaround time. The low rate of identified mutations may indicate that the majority of cases of steroid-resistant nephrotic syndrome are not due to monogenic mutations in currently known disease genes.

2573W
Diagnostic exome sequencing can alter a primary clinical diagnosis. F. Taylan1,2, M. Kvarnung1,2, A. Lindstrand1,2, T. Bu1, A. Nordgren1, E. Blennow1, M. Nordenskild1, D. Nilsson1,2, 1) Department of Molecular Medicine and Surgery and Center of Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Science For Life Laboratory (SciLifeLab), Solna, Sweden.

Whole exome sequencing (WES), which is now feasible and clinically available, allows us to identify disease causing variants in rare diseases. Advancements in bioinformatic tools also make the data produced by WES more interpretable and easy to handle. As part of our effort to use WES as a diagnostic tool, we have collected samples from distinct families with consanguineous marriages and at least two affected children with complex phenotypes. Whole exome sequencing was performed for all available family members, such as parents, affected and healthy children. The family case presented here has two affected children with a clinical diagnosis of Bardet Biedl Syndrome (BBS; MIM209900), a multisystemic ciliopathy characterized by retinal degeneration, obesity, intellectual disability and renal dysfunction. The major findings observed in the affected children were retinal dystrophy, obesity, hypogonadism, eczema, small teeth and flat feet. Whole exome sequencing revealed a mutation in any of the seventeen known BBS genes and genome wide array comparative genomic hybridization did not detect any deletion or duplication at the chromosomal level. We then expanded our analysis to other ciliopathy genes associated with phenotypes similar to BBS. A novel homozygous stopgain mutation was detected at position 2437 of the ALMS1 (ALMS1:NM_015120:exon8:c.C7310A:p.S2437X) in both of the affected children, and the mother, the only parent available, was found to be a carrier of this mutation. The ALMS1 has been known to cause Alström Syndrome, an autosomal recessive ciliopathy with a clinical picture partly overlapping with BBS. This finding demonstrates how exome sequencing can alter a clinical diagnosis and reveal differential diagnoses of phenotypically similar syndromes. As the sequencing technologies become cheaper and bioinformatic tools improve further, whole exome sequencing for clinical diagnostic purposes finds its place in routine medical practices.
2574T  
The importance of considering autosomal genes for the diagnosis of Non-Syndromic Intellectual Disability. C. Tan1, S. Topper1, V. Nelakuditi1, K. Arndt2, F. Kobiericki3, D. del Gaudio1, N. Meeks2, J. Saaril1, V. Misra3, S. Sastri4, S. Levesque5, L. Russell6, G. Sullivan7, S. Das1. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Pediatrics, Section of Genetics, University of Colorado, Aurora, CO; 3) Division of Genetics and Metabolic Disorders, Children’s Hospital of Michigan, Detroit, MI; 4) Department of Medical Genetics, Montreal General Hospital, Montreal, QC.

Intellecual Disability (ID), with a prevalence estimated to be between 1-3% of the general population, is a lifelong disability typically presenting in infancy or early childhood. It is estimated that mutations in the genes on the X chromosome may account for about 10% of all cases of ID, thus molecular investigations into the X chromosome in elucidating the etiology of an individual with ID have been routinely suggested. Due to the genetic and phenotypic heterogeneity of patients with non-syndromic ID, multi-gene testing, the concurrent analysis of multiple genetic loci, can be beneficial. The utilization of high throughput, massively parallel sequencing, or next generation sequencing (NGS), has greatly improved the ability to simultaneously analyze multiple genetic loci, and X-linked ID (XLID) panels including analysis of a large panel of X-linked genes can be used for diagnosis. We performed next generation sequencing analysis of a total of 95 genes implicated in non-syndromic ID. The 95 genes included 61 X-linked genes, 20 Autosomal Recessive genes and 14 Autosomal Dominant genes. Genes were selected based on a criterion of non-syndromic ID although some syndromic genes were also included, as a range of mutations in a single gene can sometimes confer both syndromic and non-syndromic phenotypes. As more genes are identified and implicated in non-syndromic ID, it seems appropriate to consider the inclusion of autosomal genes in multi-gene panel testing. To date, we have studied 10 patients and have identified pathogenic mutations in 3 out of the 10 patients. All three mutations are novel and include 1) G49D missense mutations in SYNGAP1 (c.1783delG and c.2602delG) and one splicing mutation in TCF4 (c.991-2A>G). All genes in which mutations were identified are autosomal, typically de-novo in inheritance. Phenotypic information was collected on all three patients. Both patients identified with mutations in TCF4 have features of microcephaly and facial dysmorphism that were possibly suggestive of a spectrum of monogenic conditions. Our three mutation-positive cases provide further evidence supporting the utility of including autosomal genes in multi-gene panel analysis of patients with intellectual disability.

2576W  
Performance analysis of saliva generated genomic DNA used for genotyping on the Affymetrix DMET Plus array as part of the Coriell Personalized Medicine Collaborative. N.P. Gentry, N.C. Werner, D.E. Lynch, L.A. Swanson, Coriell Institute for Medical Research, Camden, NJ.

The use of saliva as a source of genomic DNA for research and clinical studies has grown in popularity due to the ease of collection and participant compliance. The Coriell Personalized Medicine Collaborative (CPMC) has been using Oragene collection kits for the past six years as the source of genomic DNA, initially for genotyping on the Affymetrix Genome-Wide Human SNP Nsp/Sty 6.0 array, and for the past four years, also for genotyping on the Affymetrix DMET Plus array. Although multiple studies on the SNP 6.0 array have utilized saliva as the source of DNA, the CPMC is one of the first large scale studies to also use that DNA on the DMET Plus array. To date, the study has successfully processed more than 4000 samples on the DMET Plus array with an average call rate of 99.5%. This was similar to the 99.4% call rate achieved on the SNP 6.0 array. The genotyping results from saliva generated genomic DNA have also proven to be highly reproducible on DMET Plus. Independent extractions from individual Oragene kits as well as across multiple kits for 4 samples used as processing controls have resulted in average call rates between 99.5% and 99.8% across thirty or more replicates of each control. Furthermore, in the control replicates, the concordance rates in a set of 166 variants of interest to the CPMC study ranged between 99.7% and 99.9%. Finally, because the genomic DNA samples were run on both Affymetrix arrays, it was also possible to examine the performance of 212 SNPs that are present on both platforms. For a set of 1920 samples, the average concordance was 98.2%. When excluding SNPs where either or both arrays had a No Call, the average concordance was 99.5%. The use of saliva generated genomic DNA in the Affymetrix DMET Plus assay has proven to be very successful and has allowed the CPMC to expand its genotyping options while maintaining a single DNA source.

2577T  
Improved diagnosis of mitochondrial disorders by next generation sequencing approach. V.W. Zhang, J. Wang, Y. Feng, X. Tian, L-J. Wong. Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Background: Step-wise analyses of the genetically and clinically heterogeneous mitochondrial disorders are time consuming and cost ineffective. Next generation sequencing (NGS) technology allows simultaneous sequence analysis of any number of target genes. Method: The intact circular mitochondrial genome is enriched by long range PCR as a single amplicon and has allowed the CPMC to expand its genotyping options while maintaining a single DNA source. Independent extractions from individual Oragene kits as well as across multiple kits for 4 samples used as processing controls have resulted in average call rates between 99.5% and 99.8% across thirty or more replicates of each control. Furthermore, in the control replicates, the concordance rates in a set of 166 variants of interest to the CPMC study ranged between 99.7% and 99.9%. Finally, because the genomic DNA samples were run on both Affymetrix arrays, it was also possible to examine the performance of 212 SNPs that are present on both platforms. For a set of 1920 samples, the average concordance was 98.2%. When excluding SNPs where either or both arrays had a No Call, the average concordance was 99.5%. The use of saliva generated genomic DNA in the Affymetrix DMET Plus assay has proven to be very successful and has allowed the CPMC to expand its genotyping options while maintaining a single DNA source.

2575F  
Clinical usefulness of copy number variants detected by affymetrix high-resolution genome-wide array. E. Cho1, E. Lee2, J. Jang3, H. Kim2.
1) Green Cross Laboratories, Yongin, South Korea; 2) Gachon University, Gil Medical center, Incheon, South Korea.

Background: Recently, chromosomal microarray has proven to be an effective tool for detection of submicroscopic chromosome abnormalities causing congenital disorders and has been adopted for clinical applications. Here, we investigated the usefulness of chromosomal microarray, as a first-tier tool in detecting the etiology of developmental delay, intellectual disability, autism spectrum disorders and multiple congenital anomalies in a large number of Korean pediatric patients. Methods: We applied Affymetrix Cyto- genetics Whole Genome 2.7M array and Affymetrix Cytoscan 750K array in 333 patients (2.7M array: 200 cases and 750K: 133 cases) to assess CNV detection and evaluated the clinical significance of detected CNVs. Results: We found 103 cases (30.9%) with known pathologic CNVs (64 cases: 19.2%) and CNVs of uncertain clinical significance (39 cases: 11.7%). 3 cases (1.0%) of uniparental disomy were detected. 70% (45/64) of known pathologic CNVs were <5Mb and would likely not be detected by G-banded chromosome analysis. 50 cases with copy number losses, 45 cases with copy number gains, 6 cases with both copy number gain and loss and 2 cases with mosaic copy number gain (12q mosaic gain) were detected. The smallest pathologic CNV detected was 23kb. 2.3 Mb gain at 8p23.2 was found in 7 cases revealed as a probably benign CNV by parental study. Conclusions: We concluded microarray analysis significantly improves the diagnostic yield than G-banded chromosome analysis.
2578F Detection of Disease-Causing Mutations in the Ashkenazi Jewish Population Using a BioFilm Microarray. M. Procter¹, C. Smith², R. Mao³. 1) Research and Development, ARUP Laboratories, SLC, UT; 2) AutoGenomics, Vista, CA; 3) Department of Pathology, University of Utah School of Medicine, SLC, UT.

Background: Jews of eastern European descent, referred to as Ashkenazi Jews, are at a higher risk of carrying certain genetic diseases at greater frequencies than the general population. The American College of Obstetricians and Gynecologists (ACOG) recommends genetic testing for people of Ashkenazi Jewish descent for mutations associated with the following conditions: Tay-Sachs disease, Canavan disease, and Familial Dysautonomy. Further ACOG recommendations include testing be offered to these same individuals for mutations associated with the following diseases: Gaucher disease, Bloom syndrome, Fanconi Anemia group C, Neimann-Pick disease type A, and Mucolipidosis type IV. These autosomal recessive diseases occur 20-100 times more frequently in the Ashkenazi Jewish population, and are associated with life-threatening conditions. DNA-based carrier screens for the above conditions in this ethnic group are sensitive due to a number of common shared mutations. Materials and Methods: The assay involves multiplex PCR of DNA at 30 ng/µL, followed by analyte specific primer extension (ASPE) prior to hybridization of the ASPE primers to a BioFilm microarray. The microarray chip with bound ASPE product is washed repeatedly before optical scanning with signal detection and analysis. Steps subsequent to PCR were performed in the Infiniti™ bioanalyzer. For our evaluation, 31 samples representing all 30 mutations and 1 polymorphism were run in the assay to determine accuracy. We utilized DNA purchased from Coriell Cell Repositories and known-positive and known-negative samples from ARUP’s clinical Molecular Genetics laboratory as well as samples acquired from collaborators. Results: Twenty-six mutations were detected with 100% concordance with expected results using clinical samples, Coriell samples, and samples from collaborators. The remaining 5 variants were detected with 100% concordance with expected results using a commercially available 'supercon' containing all AJ variants included in this assay. Conclusions: We found this assay to be reliable in the detection of a large number of mutations associated with Ashkenazi Jewish diseases. The microarray format of this test will allow the detection of multiple mutations in a single reaction. Automation of all steps subsequent to the PCR step in the Infiniti™ instrument allows for high throughput with minimal hands-on time.

2579W Comprehensive mutation analysis by next generation sequencing in patients with neonatal intrahepatic cholestasis. T. Togawa, T. Sugiura, K. Ito, T. Endo, S. Saitoh. Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.

Next generation sequencing (NGS) technology has revolutionized genomic and genetic research. Neonatal intrahepatic cholestasis is a heterogeneous disorder and caused by mutations in a number of genes, making genetic diagnosis challenging. We developed a diagnostic panel of causative genes for neonatal intrahepatic cholestasis using AmpliSeq (LifeTechnologies) and the Ion Torrent PGM technology. DNA samples were obtained from 30 individuals with neonatal intrahepatic cholestasis; 4 patients with Alagille syndrome (AS), 2 patients with neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD), 4 patients with progressive familial intrahepatic cholestasis (PFIC), and 2 patients with unknown etiologies. One patient with AS, 2 patients with NICCD and 2 patients with PFIC had known mutations detected by conventional Sanger sequencing, and were invited as positive controls. A panel of genes included JAG1, NOTCH2, SLC25A13, ATP8B1, ABCB11, ABCB4, AKR1D1, HSD3B7, CYP7B1, TJP2, BAAT, EPHX1, ABC2, and VPS33B. NGS was performed on Ion Torrent PGM and variant call and annotation were performed with CLC Genomics Workbench version 6.0 (CLC bio). We detected disease-causing mutations in 6 patients: one JAG1 mutation in one patient with AS, two SLC25A13 mutations in one patient with NICCD, and 4 ABCB11 mutations in 3 patients with PFIC as well as one patient with AS. For positive controls, five single nucleotide variants (SNVs) detected by Sanger sequencing were detected by NGS, while only one out of three small insertions/deletions was identified. This study showed clinical usefulness of comprehensive mutation analysis by NGS for neonatal intrahepatic cholestasis. However, improvement on detection of small insertion/deletion remains to be fulfilled.

2580T CIGMA (Clinical Impact of Genetic Mutational Analysis): a new approach to mutational classification in large-scale clinical genetic testing. C. Turnbull¹, R. Sultana², S. Mahamdallie³, E. Ruark³, H. Han- son¹, N. Rahman¹, ². ¹) Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, United Kingdom; ²) Royal Marsden NHS Foundation Trust, London, UK.

The evolution of Next-Generation sequencing (NGS) technologies has rendered dramatic expansion in the delivery of germline genetic testing for cancer predisposition both affordable and technically viable. To deliver such large-scale expansion, germline genetic testing for cancer predisposition will likely increasingly be undertaken by oncologists non-specialist in genetics. In the current model of low volume clinical genetic testing, classification of variants represents a sizeable challenge to both those who report the tests and clinicians who manage the patients. Currently, the majority of rare variants detected are classified as ‘variants of uncertain significance’ and these are detected in 5-10% of many individual cancer gene tests. This classification can confer substantial ambiguity with regard to communication of risk and management implications. Large-scale gene panel testing will identify significant numbers of rare variants per test, potentially scaling up this ambiguity to a level rendering large-scale non-specialist delivery of these tests unfeasible. Hence, a new era of expanded genetic testing will require commensurate new approaches to classification and reporting of mutational data. Most current approaches to variant classification are largely informed by mutational data from disease cases, which enables clear classification as pathogenic or non-pathogenic for only a small minority of variants. We have undertaken large-scale analyses of mutational data from cancer predisposition genes and integrated these with case mutational data and gene-specific parameters relating to mechanisms of pathogenicity and the clinical relationship between gene and phenotype. Through large-scale bioinformatic integration of these data sources superimposed with clinical interpretation and decision-making, we have developed a novel system of variant classification to deliver clear, robust, consistent classification of mutational results into categories defined by clinical impact and management. We have developed an NGS panel comprising 97 cancer predisposition genes (illumina TruSight Cancer panel) and are piloting a new model of delivery allowing the classification of oncogenes in clinical genetic testing.


[Background and Objective] Our objective was to perform genetic analysis in Japanese patients with Charcot-Marie-Tooth disease (CMT) and identify the molecular epidemiology. A DNA microarray consisting of 28 known CMT-causative genes has been applied for mutation screening since 2004. However, the positive rate of mutation detection was as low as 10%-20%. Therefore, in May 2012, we increased the number of target genes to 60 of mutational results into categories defined by clinical impact and management. We have developed an NGS panel comprising 97 cancer predisposition genes (illumina MiSeq). [Methods] We conducted a genetic test of 182 Japanese patients with CMT from May 2012 to April 2013. Based on the median nerve motor conduction velocity (MCV), patients with CMT were broadly classified into demyelinating type (MCV < 38 m/s) and axonal type (MCV > 38 m/s). For demyelinating type CMT, the presence of the PMP22 duplication mutation was excluded before this study. When novel suspected mutations were detected, a segregation study was performed to identify the pathogenicity. The negative cases, filtered by the NGS system, proceeded to exome sequencing. [Results and Discussion] Using the NGS system, we identified pathogenic mutations in 42 (23%) of 182 patients. Thirteen mutations were identified in MFRN2 (CMTX1) and GJB1 (CMT2A2). We also discovered mutations in MPZ (6/182), NEFL (4/182), SH3TC2 (2/182), GDAP1 (2/182), PRPS1 (1/182), and TRPV4 (1/182). However, segregation study data were not available for 20 variants, although they were suspected as causative mutations. If we considered these variants, the positive rate would be improved to 34% with the NGS system.
2585W
Multiplex-PCR coupled to Next-Generation Sequencing (NGS) and SNP array technologies greatly improve molecular diagnosis of Usher syndrome. C. Bonnet1,2, S. Chantot-Bastarda2, I. Slesioraty1,2, A. Fakiri1, F. Testa1, L. Martorell Sampoli1, S. Ghebri1, S. Dad1, S. Marlin1,2, S. Koh1, D. Zobor1, S. Mohand-Said1,2, F. Simonelli6, S. Banfi6, J. Rodriguez Jorge7,1, J. L. Birken Moller4, A. Kurtenbach2,4, M. Hawlina6, A. Aurichio6, J. A. Sahel1,2,1A. Tavol1,2,6, E. Zrenner2, C. Pet1,2,9,1,1.

Usher syndrome (USH) is the most prevalent cause of hereditary deafness-blindness in humans. There are three types of USH (USH1, USH2, and USH3) have been distinguished clinically. These are defined according to the severity of the sensorineural hearing impairment, the presence or absence of vestibular defects, and the precocity of retinitis pigmentosa onset. To date, 12 USH loci have been characterized and ten causative genes identified. In order to improve molecular diagnosis and to minimize its cost, we here developed a multiplex-PCR coupled to Next-Generation Sequencing (NGS) technologies. By using this unique approach in a large cohort of 400 diagnosed USH patients, biallelic mutations were detected in 83% of patients, monoallelic mutations were detected in 13% of them and no mutation was found in 4% of patients. SNP array was performed for patients with no mutation or with concomitant clinical and genetic data suggestive of abnormal mitotic segregation (de novo mutation and/or duplication) were observed in 25% of the USH patients with only one pathogenic mutation. The patients with no mutation found, were examined by whole exome sequencing. Of note, one patient carried a homozygous nonsense mutation in USH2A and one pathogenic nonsense mutation in USH1G. Another patient carried two pathogenic mutations (nonsense and deletion) in USH2A and two predicted pathogenic mutation (splice site variation and predicted pathogenic missense mutation) in CDH23. These observations are crucial in the perspective of gene therapy. An early and reliable diagnostic test is one critical step to develop specific and adapted therapies to cure this disorder.

2585F
Next-Generation Sequencing in the Molecular Diagnostics of Rare Diseases using a Gene Panel Approach. S.H. Eck, I. Vogl, S. Datser, S. Kuecuek, W. Rupprecht, B. Busse, J. Hoefele, S. Chahrokh-Zadeh, C. Marschall, K. Mayer, I. Rost, HG. Klein. Molecular Genetics, Center for Human Genetics and Laboratory Diagnostics, Dr. Klein, Dr. Rost, Martinsried, Germany.

The advent of Next-Generation Sequencing (NGS) in clinical diagnostics opens vast opportunities through the ability to simultaneously sequence all genes contributing to a certain disease at lower cost and higher speed compared to traditional sequencing approaches. In rare and heterogeneous disorders, NGS may lead to a significant improvement of the diagnostic yield. On the other hand, the practical implementation of NGS in a clinical diagnostic setting involves a variety of new challenges which need to be overcome. Among these are the generation, analysis and storage of unprecedented amounts of data, strict control of sequencing performance, validation of results, interpretation of detected variants and reporting. Here we present a panel approach for the molecular diagnostics of rare disorders. Exonic regions of more than 250 custom selected genes are enriched in parallel by oligonucleotide hybridization and capture (Illumina TSC) and sequenced on the Illumina MiSeq instrument. During analysis, only genes from the requested indication (grouped in subpanels) are selected to limit analysis to relevant genes, while simultaneously minimizing the risk of incidental findings. Data analysis is performed using the CLC Genomics Workbench (v. 6.0.3, CLCbio) and custom developed Perl scripts. Target regions which fail to reach the designated coverage threshold of 20X are reanalyzed by Sanger sequencing and candidate mutations are independently confirmed in a separate reaction. All detected variants are imported into an in-house created database. Remapping of reads against the relevant region which can be web interactively randomly shown using dynamic data analysis and filtering. Information from all 250 genes is used in an anonymized way for internal variant frequency calculation, quality control and the detection of potential sequencing artifacts. We have applied this approach to more than 150 samples from a variety of different disorders. In particular we use the outlined approach for the diagnostics of arthrogryphic cardiac disorders (LQTS, HCM, DCM), connective tissue disorders (EDS, TAAAD), rare kidney disorders (Nephrotic Syndrome, CAVUT), neurodevelopmental disorders (Nörgen syndrome, Microcephalies), metabolic disorders (MODY diabetes) and pharmacogenetics.

2583T

Hearing impairments can be classified in many different ways and have an incidence rate of approximately one in 1000 births and additionally affect 50% of the population at age 80 or higher. There are multiple genetic or non-genetic causes for hearing loss. The non-genetic factors include ototoxic drugs, perinatal infections or traumas. In most cases both exogenous factors and mutations in one or more genes contribute to the phenotype. 80% of all familial, non-syndromic deafness cases are inherited in an autosomal manner. 25% of familial deafness (autosomal dominant, autosomal recessive, X-linked recessive or maternal allele transmission) are responsible for the phenotype. Sometimes heterozygous mutations in GJB2 (gap junction protein connexin 26) are responsible for the phenotype. Sometimes heterozygous mutations in GJB2 can occur in combination with a heterozygous deletion del(GJB6-D13S1830) in GJB6 (gap junction protein connexin 30). Furthermore there are over 70 genes known which can cause different types of deafness (autosomal dominant, autosomal recessive, X-linked recessive or mitochondrial). In a pilot study we designed a deafness gene panel comprising 75 nuclear genes and 6 mitochondrial genes associated with hearing impairment. In addition to 3 positive and 7 negative controls we sequenced 2 related patients with non-syndromic deafness. Conventional techniques and Sanger sequencing and GJB2 sequencing and GJB6 deletion analysis was exhausted and yielded no results. Sequencing was performed on the Illumina MiSeq Next-Generation Sequencing platform. Data analysis was performed using CLCbio workbench (v6.0.7) and custom Perl scripts. The target regions, in total encompass 90,135,454bp and 16,569bp were selected in parallel by oligonucleotide hybridization and capture (Illumina TSC). On average, 40-70% of the reads could be mapped to the human genome (build hg19) and 30-60% to the mitochondrial genome, of which between 55-60% and 96-97% respectively were considered on target. The target regions were selected on the basis of the above mentioned controls a variant calling pipeline was established and validated. Using this pipeline the 2 patients (mother and daughter) were analyzed and the AT4745G mutation of the mitochondrial genome was detected. This mutation is known as a likely pathogenic variant and was confirmed in the 2 related patients by Sanger sequencing for both patients independently. Any intake of ototoxic drug was not investigated. The mutation was confirmed independently by Sanger sequencing.

Introduction: Fetal cell-free DNA (cfDNA) in maternal plasma enables techniques for screening for fetal aneuploidies using next generation sequencing technologies. It is known that the power to detect fetal aneuploidy is directly related to the amount of cfDNA, regardless of whether the approach employs sequencing random DNA fragments using a shotgun method or a directed analysis of targeted DNA fragments. Here, we demonstrate that Digital Analysis of Selected Regions (DANSR™) assays for single-nucleotide polymorphisms (SNPs) on chromosomes 1-12 and chromosome Y accurately and reproducibly measure the fraction of cfDNA in maternal plasma (fetal fraction). We present data showing fetal fraction measurements using SNPs or chromosome Y alone are highly correlated, and these measurements are reproducible, as evidenced using two blood tubes from a single blood draw. Objective: To report clinical reproducibility results on Harmony Prenatal Test's ability to measure fetal fraction. Methods: This study included 11,024 maternal plasma samples containing observable counts on the Y chromosome. Samples were processed between 15 March 2013 and 17 May 2013 from pregnant women carrying a fetus at least 10 weeks in gestational age. A separate set of 1,266 research samples, in which both blood tubes were drawn at the same time but processed on different dates, were used to test the reproducibility of the fetal fraction estimates. Samples were processed as previously described (Norton et al., 2012). Fetal fraction was assessed simultaneously by assays against a set of 192 SNP loci on chromosomes 1-12 (FF-SNP) and a set of Y assays (FF-Y). Results: In the first study of 11,024 samples, the FF-SNP and FF-Y were correlated with an r-squared of 0.99. Among the duplicated blood tubes, the median processed time difference between a pair of blood tubes was 80 calendar days. Despite this large time difference and various possible reagent, robotic, and operator differences, the fetal fraction calculated between paired samples was r-squared 0.95, a slope of 1. The median absolute relative difference between fetal fractions measured by the tube pairs was 4.5% (S.E. 0.2%). This implies a fetal fraction of 0.04, on average, may result in either 0.038 or 0.042 on a repeated measure from a different blood tube.

Conclusions: Fetal Fraction estimates from the Harmony Prenatal Test are accurate, precise, and reproducible.


By recent progress in NGS technologies, a genomic era for cancer studies is growing rapidly. In this era, automated capillary electrophoresis (CE) DNA sequencer by sanger dideoxy terminator is still a widely used for validating gene sequences, and is considered gold standard for target gene Analysis. However, CE DNA sequencers cannot detect low-frequency mutation, because sequencing workflow and subsequent signal processing of CE DNA sequencers are optimized for detecting germline mutation (50% existence). To address this issue, we developed a new method for detecting mutation with high sensitivity using CE DNA sequencers. In the method, we first modified sequencing workflow, namely, measuring four kinds of bases in separate capillaries, to reduce to cross-talk-fluorescence signal. Second, we implemented a procedure of extracting relative intensity of each peak to a neighbor peak of the same kind of base in a raw fluorescent signal data. Dispersion of relative intensity of each peak was found to be small enough to detect mutation with high sensitivity by comparing with relative intensity of reference data. Third, we developed method successfully detected 5% existence mutation in EGFR and KRAS without visual inspection. The proposed method shows a possibility to detect mutation of much lower existence, and is promising for practical use to detect genetic mutations.

2587F Fast STR-PCR protocol enabling rapid and high quality chimerism analysis after allo-HSCT. W. Teng1, H. Liu1, F. Wang1, Y. Wang1, X. Chen1, J. Fan1, P. Zhu1. 1) Molecular Medicine Lab., Hebei Yandai Hospital, Sanhe, Hebei, China; 2) Department of Hematology, Peking University First Hospital, Beijing, China.

Background: Chimerism analysis based on STR-PCR is the standard method for chimerism analysis after allogeneic hematopoietic stem cell transplant (allo-HSCT). The cycling time needs about 3.5 hours, and incomplete adenylation of PCR products often become a problem. Here we aimed to develop a fast PCR protocol enables rapid and high quality chimerism analysis. Methods: Peripheral blood and fingernail samples were collected from healthy donors and patients. Simulated mixed chimerism samples were prepared by fold dilution. The AB Identifiler Kit with 15 STR loci was used. Primer: All the four kinds of polymerase can be successfully used for STR-PCR and genotyping. The STR-PCR cycling time for MyTaq, Phusion and Q5 were about 30 minutes, considerably less than 3.5 hours when using AmpTaq Gold. For the 3 kinds of fast polymerases, the Q5 polymerase showed the best balanced amplification efficiency among the 15 STR loci. The PCR product of Phusion and Q5 don't have non-template addition ‘A’ tail due to lack of adenylate activity, yet the allelic ladders without ‘A’ tail must be specially made. The MyTaq has extremely efficient of adenylate activity and showed complete adenylate activity yet the allelic ladders without ‘A’ tail must be specially made. The MyTaq has extremely efficient of adenylate activity and showed complete adenylate activity. We have used a split read algorithm to detect the inversion from 14% coverage from 1.72 billion reads, finding a fetal DNA composition of ~25% of reads. We used a split read algorithm to detect the inversion from 14% of reads (~28% fetal coverage at the locus), demonstrating feasibility of SV detection from very deep WGS of cfDNA. Transcriptome studies are ongoing. In the twins (DGAP258) we identified a pericentric inversion that rearranged the known copy number morbid 6p25.3 locus, as well as an independent cryptic inversion, but neither disrupted genetic sequence. The consequences of inverting a recurrent copy number morbid region but not disrupting genetic sequence is unknown, but the twins were recently born with a well newborn examination. These data highlight the power of prenatal WGS by jumping libraries to detect pathogenic SVs of all classes, however they are reliant on invasive methods that carry substantial risk to the fetus. They also show the potential feasibility of SV detection by non-invasive methods, though codes a clinical effort far from oxymoronic. Alternatively, our studies emphasize the unique challenges facing development of a comprehensive prenatal genetic screening strategy that accesses the entire pathogenic mutational spectrum but minimizes risks to the fetus.
2590F
A novel and rapid digital PCR-based method for the identification of 22q11.2 Deletion Syndrome in large population screening. V. Hwang1, D. Maar2, J. Regan2, T. Simon2, F. Tasco2. 1) Department of Biochemistry and Molecular Medicine, UC Davis, Davis, CA; 2) Digital Biology Center, Bio-Rad Laboratories, Pleasanton, CA; 3) MIND Institute, UC Davis Medical Center, Sacramento, CA.

22q11.2 Deletion Syndrome (22q11DS), the most common microdeletion syndrome in humans, is characterized by a wide range of clinical manifestations including craniofacial defects, developmental delay, neurological and psychiatric problems, autism spectrum disorders and congenital heart defects. 70-80% of individuals have a 3 Mb deletion, 15-30% have a nested 1.5 Mb deletion, and the remainder have atypical deletions within the 22q region. Because the syndrome is associated with over 100 different diagnostic findings that range from mild to severe and life-threatening conditions, diagnosis is often delayed or missed. Therefore, due to this high variability, propositions to include 22q11DS in newborn screening panels have been suggested.

The currently accepted method for diagnosis is fluorescent in-situ hybridization (FISH). This approach, however, is expensive, labor intensive, and requires special technique and equipment. We have developed an inexpensive, rapid, sensitive and specific test that can identify newborns with 22q11DS. The approach uses a droplet digital PCR (ddPCR) that processes samples collected on blood spot cards and provides accurate and rapid diagnosis while keeping costs at a minimum-the ideal premise for large population screenings, such as newborn screening. We performed a validation study where we blindly screened over 500 anonymous blood spots collected from the general population that included a small number of 22q11DS patients. We were able to detect 93% of deletions using our ddPCR method. The median threshold that ddPCR could detect was 0.01%. We saw 97% sensitivity and >99% specificity to detect single exon CNVs. Thus, diagnostic testing laboratories often resort to insilico genotyping the genomic location of the endpoints of copy number variation (CNV) to improve patient access to comprehensive genetic testing.

Materials & Methods: 137 samples (blood and cell-line derived DNA samples) were selected to maximize variation across the genome. The CytoScan Dx assay was run at Affymetrix and next generation whole genome sequencing was independently performed and analyzed in a blinded fashion at Complete Genomics. Endpoint agreement was pre-specified as the sequence-defined endpoint of ± 12 markers for loss CNVs and ± 25 markers for gain CNVs, translating into an effective resolution of 25kb for losses and 50kb for gains. In addition, copy number accuracy was assessed by comparing results with those determined by routine patient care methods such as FISH or karyotype.

Results: The overall endpoint agreement was 93.4% and was similar for copy number loss regions (92.5%) and for copy number gain regions (94%). No cases were missed by the current algorithm. Additional validation results will be represented, under a simulation model, altering the observed laboratory data as a function of error. We achieved a sensitivity and specificity of >99% to detect single exon CNVs, translating into an effective resolution of 25kb for losses and 50kb for gains. In addition, copy number accuracy was assessed by comparing results with those determined by routine patient care methods agreed 100% of the time on copy number variation.

Conclusions: The CytoScan Dx assay is highly accurate in detecting copy number variation. For those copy number variation regions of 25kb and 25 markers for loss segments and 50 markers and 50kb for gain segments, CytoScan Dx assay was able to determine the genomic location within ± 1 marker of the sequence determined endpoint demonstrating very high resolution and a high degree of accuracy in determining the copy number variation endpoint locations.

2592T
A comparison of CNV endpoint accuracy between CytoScan®Dx assay and Next Generation Sequencing. A. Roter1, B. Eynon2, S. Close2, K. Kwiatkowski1, D. Ballinger2, S. Yang1, R. Duttagupta1, C. Chen2, K. Sugeng1, A. Singh1, T. Chen1, M. Chadha1, E. Fung1. 1) Affymetrix, Inc., Santa Clara, CA; 2) Indiana University School of Medicine, Division of Clinical Pharmacology, Indianapolis, IN; 3) Complete Genomics, Mountain View, CA.

Background: The genome-wide distribution of ~2.7 million markers enables CytoScan®Dx assay to determine copy number state with very high resolution. To assess the resolution of CytoScan Dx assay, the accuracy in determining the genomic location of the endpoints of copy number variation (CNV) was evaluated. Endpoint evaluation was compared between next generation sequencing and CytoScan Dx assay for 1,387 concordant CNV endpoints in post-natal samples.

Materials & Methods: 137 samples (blood and cell-line derived DNA samples) were selected to maximize variation across the genome. The CytoScan Dx assay was run at Affymetrix and next generation whole genome sequencing was independently performed and analyzed in a blinded fashion at Complete Genomics. Endpoint agreement was pre-specified as the sequence-defined endpoint of ± 12 markers for loss CNVs and ± 25 markers for gain CNVs, translating into an effective resolution of 25kb for losses and 50kb for gains. In addition, copy number accuracy was assessed by comparing results with those determined by routine patient care methods such as FISH or karyotype.

Results: The overall endpoint agreement was 93.4% and was similar for copy number loss regions (92.5%) and for copy number gain regions (94%). No cases were missed by the current algorithm. Additional validation results will be represented, under a simulation model, altering the observed laboratory data as a function of error. We achieved a sensitivity and specificity of >99% to detect single exon CNVs, translating into an effective resolution of 25kb for losses and 50kb for gains. In addition, copy number accuracy was assessed by comparing results with those determined by routine patient care methods agreed 100% of the time on copy number variation.

Conclusions: The CytoScan Dx assay is highly accurate in detecting copy number variation. For those copy number variation regions of 25kb and 25 markers for loss segments and 50 markers and 50kb for gain segments, CytoScan Dx assay was able to determine the genomic location within ± 1 marker of the sequence determined endpoint demonstrating very high resolution and a high degree of accuracy in determining the copy number variation endpoint locations.

2591W
Detection of 22q11.2 deletion syndrome in Colombian patients with isolated congenital cardiopathy by MLPA. T. PINEDA1, O. MORENO2, I. ZARANTE1. 1) Medical Genetics Resident, Genetics Institute, Pontificia Universidad Javeriana, Bogotá, Colombia; 2) Biol. Msc. Genetics Institute, Pontificia Universidad Javeriana, Bogotá, Colombia.

22q11.2 deletion syndrome is one of the most common genetic syndromes and it has a wide phenotypic spectrum, abnormalities can include cardiac defects, immunodeficiency, hypocalcaemia, speech and language impairment, cleft palate, mental retardation and any other major anomaly. The results showed that the classical 22q11.2 deletion is present in the 15.3% of the patients (4/26), we did not find any atypical deletion in this group. With this study we can conclude that MLPA is a very useful molecular method that provide an accurate diagnosis and that it should be implemented in all patients presenting with isolated congenital cardiopathy related to 22q11.2 deletion syndrome.

2593F

Germline copy number variants (CNVs) can be detected from next-generation sequencing (NGS) data generated using targeted DNA capture technologies (e.g. exomes and other panels), however methods for doing so must overcome many technical challenges. Several algorithms have been published to detect CNVs in such data, though they may not yet be adequate for use in diagnostic testing laboratories, particularly for detection of small single-exon CNVs. Thus, diagnostic testing laboratories often resort to expensive and low-throughput methods such as MLPA to discover and confirm small CNVs. As a result, clinicians must carefully decide whether to order both a sequencing test and a deletion/duplication test for their patients. A single test that can accurately assay both types of alterations would improve patient access to comprehensive genetic testing.

We present a new method, CNVitae, which is designed to detect single-exon CNVs as well as larger regions sequenced using NGS. CNVitae is based on a statistical model for read counts and employs model-based segmentation algorithms optimized for use with sparsely distributed and high noise targets across the genome. This framework estimates the most likely copy number for all segments, and, critically for clinical use, each called segment is assigned a robust quality score indicating confidence in the copy number determination. We evaluated CNVitae on deep depth targeted NGS data generating using Agilent SureSelect capture and Illumina MiSeq 2x150 paired-end sequencing. Under a simulation model, altering the observed laboratory data in silico, we achieved a sensitivity and specificity of >99% to detect single exon hemizygous deletions at a strict confidence threshold of 25Q (probability of error <0.5%). We saw 97% sensitivity and >99% specificity to detect single exon duplications (CN=3) while four exon duplications were detected with sensitivity of >99%. In 40 patients independently known to carry clinically relevant CNVs, we detected all 37 of the single-exon or larger events with high confidence, although sensitivity to detect single exon CNVs was not detected by the current algorithm. Additional validation results will be represented, including results on additional clinical samples, the Get-RM Copy Number Variation Reference Panel, and a re-analysis of a subset of exome data from the 1000 Genomes Project.
2594W
Assessing Common Maternal Copy Number Variation during cfDNA Analysis for Non-Invasive Prenatal Testing using Digital Analysis of Selected Regions (DANSR™) Assays
C. Struble, E. Wang, J. Schmidt, A. Batye, T. Muscì, K. Song, A. Oliphant, Ariosa Diagnostics, Inc. San Jose, CA.
Objective: To report on Harmony™ Prenatal Test’s ability to provide accurate results in the presence of a maternal CNV. As previously described, DANSR assays were used to analyze common trisomies in maternal samples. The DANSR assays were selected to target nucleic acid regions unique to the chromosomes of interest in a maternal sample, and to avoid common copy number variations (CNVs) (Sparks et al., 2012). Fetal fraction was assessed in the maternal samples simultaneously with the trisomy analysis using DANSR assays against a set of 192 single nucleotide polymorphism (SNP)-containing loci on chromosomes 1-12. Using the estimated fetal fraction, a sample specific expectation is determined to evaluate if increased chromosomal counts using the DANSR assay are consistent with fetal aneuploidy and/or other clinical features such as maternal CNV. Results: In approximately 64,000 samples, 35 (0.05%) putative maternal CNVs were observed. Sixteen of these 35 (47%) belong to a 1.5 mega base maternal duplication on chromosome 21q21.3. Of these 16 cases, there was one case of a high-risk for trisomy 21 result, and 15 cases of low-risk results in the presence of this maternal CNV. The rate of high-risk for trisomy 21 was subsequently confirmed by amniocentesis. Genes in this duplicated region includes two secreted metalloproteases, ADAM metalloepitidase with thrombospondin type 1 motif (ADAMTS1) and ADAM metallopeptidase with thrombospondin type 5 motif (ADAMT5). ADAMTS1 has been reported to play a critical role in follicular rupture. ADAMTS5 has been reported to degrade aggrecan, which is a proteoglycan of the cartilage and is associated with inflammation and arthritis. An additional gene in this region include the Cysteine and proline-rich 1 (CYPR1) with high conservation among vertebrates and high expression in cells belonging to the diffuse neuroendocrine system. Conclusion: DANSR’s design allows for efficient and uniform sequencing across the chromosome loci with enough depth to evaluate the risk for fetal aneuploidy in the presence of previously unreported maternal CNV.

2596V
Angelman and Rett syndromes are common genetic disorders that share many similar features, including developmental delay, intellectual disability and seizures. The genetic heterogeneity among these two syndromes and many other similar disorders make it challenging to determine the appropriate second tier of diagnostic testing since the most common causes have been ruled out. In an effort to provide a more rapid, comprehensive and cost-effective analysis for patients with features overlapping those of Rett and Angelman syndromes, the Molecular Diagnostic Laboratory at the Greenwood Genetic Center (GGC) has designed and validated a 19-gene panel which utilizes RainDance™ Technologies microdroplet enrichment and SOLiD™ Next Generation Sequencing (NGS). The panel includes MECP2 and UB3A along with 17 additional genes known to cause conditions that often present with clinical features similar to Rett and Angelman syndromes. The coverage of the panel includes all exons for each gene as well as flanking intronic regions. Thus far, 38 samples have been submitted for clinical testing. Of the 17 completed analyses, eight patients had normal results, and six patients had at least one variant of unknown clinical significance. We identified pathogenic changes in three patients (~18%). Interestingly, one of these patients was found to have a one base pair deletion in MECP2 (c.271delC). MECP2 sequencing had not been previously requested for this patient since she does not have a classic Rett phenotype. Unlike most patients with Rett syndrome who have microcephaly, short stature and cachexia, at seven years of age this patient has a normal head circumference (75th percentile), height (90th percentile) and weight (75th - 80th percentile). She is ambulatory and non-verbal with severe intellectual disability. Therefore, this panel may also be useful in detecting MECP2 and possibly UB3A mutations in individuals with atypical presentations of Rett and Angelman syndromes, in addition to being a second tier test for MECP2 and UBE3A mutation-negative patients. Our initial findings indicate that this expanded Rett/Angelman panel will provide a cost-effective method for testing the genes believed to be of the highest diagnostic priority for individuals who fall into this clinical spectrum.

2595T
Using next-generation sequencing for clinical diagnostic testing poses significant challenges associated with analytical sensitivity and specificity. Discovery of false positive variants lead to misdiagnoses and false negatives neglect clinically relevant ones. Understanding the properties of false variant calls from next-generation sequencing platforms will assist with tailoring the variant calling programs to generate high confidence sequence variants important for clinical diagnostics. The molecular genetics laboratory at the Hospital for Sick Children in Toronto, Canada is developing next-generation sequencing panels using the Ion Torrent platform. To observe the properties of false variant calls, three separate runs were performed using a known genotyped sample - J. Craig Venter - with the Inherited Disease Panel from Life Technologies. The Inherited Disease Panel targets the coding regions of over 300 genes associated with human genetic conditions and covers a wide variety of sequencing contexts that is important in understanding the properties of variant calling and their genotypes. After sequencing on the Ion Torrent platform, the data was analyzed to determine the copy number status (hg19) using TMAP 3.4.3 and variant calls were made using the Ion Torrent Variant Caller’s (3.4.2) default settings with an average read length of 135 base pairs. Of the known 1,005 SNPs present in the area targeted from the Venter sample an average of 74 variants were called incorrectly (93% sensitivity) and of the 150399 known references bases sequenced 74 were missed (99.9% specificity). The average read depth for correct homozygous calls was 347 (s.d.=312) versus 157 (s.d.=173) for false positives and for true heterozygous calls the read depth was 374 (s.d.=325) versus 312 (s.d.=230) for correct and for false heterozygous calls. The ratio of the alternative allele was 1.00 (s.d.=0.02) for correct homozygous calls and 0.81 (s.d.=0.15) for false homozygous calls. Average heterozygous alternative allele ratio was 0.49 (s.d.=0.07) for correct calls and 0.30 (s.d.=0.27) for false positives. There were no significant differences in strand bias. Limiting the analysis to read depths over 350 improved the positive predictive value from 93% to 96%. By designing testing strategies that generate read depths above 350 and narrowing the filtering range of the allelic ratios we can improve the confidence of the variant calls important for clinical diagnostics.

2597W
High-throughput screening for SMN1 copy number loss by next-gener- ation sequencing. E. Boyden, G. Porreca, M. Umbarger. Good Start Genetics, Cambridge, MA.
Spinal muscular atrophy (SMA) is a lethal autosomal recessive neuromuscular disorder caused by functional loss of the SMN1 gene. The high carrier rate of ~1 in 40 is attributable primarily to SMN1 copy number loss produced by either deletion of part or all of SMN1, or conversion of SMN1 to SMN2, a linked paralog that encodes an identical protein but is poorly expressed due to a silent non-coding variant that disrupts proper splicing. Clinical SMA carrier screening is currently performed via methods such as multiplex ligation-dependent probe amplification (MLPA) that enable the assessment of the copy number status of SMN1. Our method shows sensitivity and specificity for detection of SMN1 copy number loss similar to that of MLPA, but has the advantage of being compatible with automated high-throughput screening. Furthermore, because our method uses sequencing as the read-out, it can detect deleterious SMN1 point mutations and indels that would be missed by MLPA and related approaches. Thus, our new SMA carrier screening assay represents a significant improvement over standard SMA carrier screening methods.
2598T
Establishing Performance Specifications for Clinical Whole Exome Sequencing. M.O. Dorschner1,2, S.J. Anover-Somcke3, J. Gasper3, S. McGee4, T. Shaffer2, K. Patterson2, J.D. Smith2, G.P. Jarvis2,3, D.A. Nicker-son2, 1) Psychiatry & Behavioral Sciences; 2) Genome Sciences; 3) Depart-ment of Medicine, Division of Medical Genetics, University of Washington, Seattle WA.
Whole exome sequencing (WES) is rapidly becoming an effective tool for the molecular diagnosis of rare genetic disorders. WES is particularly well suited for disorders in which the genes are known but diagnostic assays are unavailable or those conditions with closely, difficult to distinguish phenotypes. Prior to releasing WES as a clinical test, our laboratory evaluated performance specifications in accordance with regulations set forth by the Clinical Laboratory Improvement Amendments (CLIA) or College of Ameri-can Pathologists (CAP). Sanger-based sequencing tests have typically been validated by the examination of a small set of known, mutation-positive and negative samples. The ability to detect these variants accurately, serve as the basis for analytical sensitivity, specificity and reproducibility of the test. Testing only a small number of variants, from a limited set of samples is not adequate to represent the performance of large-scale assays such as WES and whole genome sequencing. To validate our WES protocol, we chose to examine 12 well-characterized HapMap samples, in duplicate. All samples were previously genotyped using the Illumina 1M array, and >80,000 markers overlapping the exome were examined for concordance with exome sequence data. With a mean coverage of 120X (90% of the exome covered at >20X), we calculated the sensitivity and accuracy of our WES protocol at >99% for single nucleotide substitutions. Poorly covered exons were largely the result of high GC content and/or low read mapability. We believe that combining these two factors in a much higher variance reflects the true performance characteristics of our WES assay. With such high accuracy, we have developed an algorithm based on coverage and sequence quality to eliminate the need for Sanger confirmation. Only those variants exhibiting sub-optimally, coverage, quality or strand bias are routinely Sanger confirmed. Sequencing technology continues to evolve at a rapid pace necessitating an organized procedure for re-validation of WES proto-cols as changes are incorporated into clinical workflows. Scalable quality control measures will be required for wider implementation of genomic medi-cine.

2599F
Development and validation of a synthetic, single-reagent, positive control for comprehensive high-throughput carrier screening. A.M. Fedick1,2, C. Jalas3, N.R. Treff2,2, 1) Department of Microbiology and Molec-ular Genetics, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ; 2) Reproductive Medicine Associates of New Jersey, Basking Ridge, NJ; 3) Bonel Olam, Center for Rare Jewish Genetic Disorders, Brooklyn, NY.
Carrier screening tests that offer parallel analysis of multiple mutations, genes, and diseases are in high demand. Contemporary platforms include testing for hundreds of pan-ethnic mutations at once using methods such as TaqMan allelic discrimination. DNA microarrays, ordered, and pooled together, were validated on a 7900 HT real-time PCR instrument using 384 well plates. Two synthetic controls were designed to amplify and were excluded from further analysis. The remaining controls were added at a 1:1 ratio with DNA from an individual known to possess novel alleles for all mutation sites. These constructs were used in conventional TaqMan allelic discrimination analysis using first a 7900 and then a QuantStudio 12K Flex Real-Time PCR System for high-throughput genotyping. Successful genotypes were observed for 98.1% (26/27) of the 12 genes, for 99.0% and for 92.6% (25/27) on the QuantStudio 12K Flex. The two failed controls both involved mutations in the GBA gene, which has a known pseudogene, and are now being redesigned to achieve 100% coverage. In conclusion, this proof-of-principle study demonstrates a novel strategy to provide a single reagent suitable for routine quality control for high throughput carrier screening.

2600W
Limited clinical utility of Whole Exome Sequencing in the diagnosis of hereditary neuropathies. O. Jarrova1,2, J. Warman1,2, J. Schwartz-tenbrueker1, C. Goldsmith1, N. Carlson1, E. McCready3, G. Yoon3, S. Bakers5, A.M. Innes6, C. Beaulieu1, A. Smith1, T. Hartley1, K. Boycott1, 1) Children's Hospital of Eastern Ontario, Ottawa; 2) Division of Neurology, Department of Medicine, The Ottawa Hospital; 3) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec; 4) Division of Clinical & Metabolic Genetics, The Hospital for Sick Children, Toronto; 5) Department of Medi-cine, Divisions of Physical Medicine and Rehabilitation, McMaster University, Hamilton; 6) Department of Medical Genetics, Faculty of Medicine, University of Calgary, Alberta.
Hereditary neuropathies comprise a distinct group of inherited disorders in which peripheral nerves undergo progressive degeneration, leading to significant physical disability of the hands and lower legs. While autosomal recessive and X-linked forms have been described, hereditary neuropathies predominantly show autosomal dominant inheritance. Causative mutations can reside in any of a number of genes responsible for various aspects of maintenance of peripheral nerves and result in an overlapping clinical spectrum for these disorders. Currently available molecular diagnostic options only allow the sequential analysis of a limited number of these genes. Therefore determining the genetic cause for a specific patient/family is often time-consuming and expensive.
We questioned whether Whole Exome Sequencing (WES) could facilitate the identification of causative mutations in patients with hereditary neuropa-thies. We sequenced the exomes of twelve unrelated patients without molec-ular diagnoses after mutations in the four common neuropathy genes (PMP22, GJB1, MPZ, MPN2), which account for over 70% of molecular diagnoses, had been ruled out. We screened the sequence data for the presence of likely damaging coding and splice-site variants using an in-house analysis pipeline and manual review of variants. While this approach has proven to be successful for diagnosis of other genetic disorders, it failed to identify obvious causative mutations in our patients. Our findings suggest that at this time using WES for the diagnosis of patients with hereditary neuropathies is challenging. This study emphasizes the need for refinement of analysis methods and further assessment of clinical utility of WES for the diagnosis of autosomal dominant heterogeneous conditions.

2601T
Comprehensive massive parallel DNA sequencing strategy for the genetic diagnosis of the Neuro-cardio-facio-cutaneous syndromes. A. Justino1,2, P. Dias1, M.J. Pina1, C. Ribeiro1, S. Sousa1, L. Ornes3, A.B. Sousa1, J.L. Costa1, J.C. Machado1,2, 1) Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; 2) Department of Genetics, Hospital de Santa Maria de Lisboa, Portugal; 3) Medical Faculty of the University of Porto, Porto, Portugal.
Background: The Noonan, Cardio-facio-cutaneous, Costello and LEO-P-ARD syndromes are members of the Neuro-cardio-facio-cutaneous syn-dromes group (NCFCS). Mutations in 11 genes have been causally linked to these disorders (PTPN11, SOS1, RAF1, RAS, MEK1, MEK2, NRAS, KRAS, HRAS, SHOC2 and CBL). Recently, an exome sequencing study also associated the gene A2ML1 with this syndrome. Due to the genetic and clinical heterogeneity of these disorders it is challenging to define straightforward strategies of sequential genetic analysis for their molecular diagnosis. Therefore, the aim of this study was to develop and validate a massive parallel sequencing (MPS) based strategy for the molecular diagno-sis of NCFCS. Methods: A multiplex PCR-based strategy for the enrichment of 11 genes on Illumina sequencing platforms (Illumina HiSeq) was developed. Two sets of genomic DNA samples of clinically defined cases of NCFCS were studied using the Ion PGM: a training set (15 cases) used to optimize the strategy and a validation set (20 cases) used to validate and evaluate the power of the new methodology. Sanger sequencing was performed to confirm all variants and fill in regions in insufficient coverage. Results: All variants identified by Sanger sequencing were detected with our MPS approach. The most frequent mutated gene was PTPN11 (n=9). Other known disease causing mutations were found in RAF1 (n=1), SHOC2 (n=1) and RAF1 (n=2). Additionally, 10 genetic variations of unknown significance were identified in SOS1 (n=2), CBL (n=3) and the novel A2ML1 (n=10). The methodology resulted in an experimental approach with a specificity of 99.4% and a maximum analytical sensitivity ≥98.1% with a power of 100% of the 12 genes increased in 15% diagnostic yield of the strategy currently used by our laboratory for the molecular diagnosis of NCFCS. Conclusions: Here we present a work-flow that provides a comprehensive genetic screening strategy for patients with NCFCS. Additionally, our approach demonstrates the potential of a combined MPS-Sanger sequencing based strategy as an effective diagnostic tool for heterogeneous diseases.

Posters: Clinical Genetic Testing

1) Life Technologies, South San Francisco, CA; 2) American Red Cross, Philadelphia, PA; 3) Diatherix Laboratories, Inc, Huntsville, AL.

Transfusion patients depend on well characterized blood donations. Incompatible blood types can cause severe hemolytic reactions. Blood group antigens have a genetic basis, in most instances they are based on a single nucleotide polymorphism (SNP). Genotyping is more accurate than serologic based tests since polyclonal antibodies may miss detection of low expression of antigens. For some blood group antigens, no antibodies are available at all. We developed a high-throughput screening panel on the TaqMan® OpenArray® covering 28 markers across 14 blood groups: Colton Co(a), Cromer Cr(a), Diego Di(a/b), Dombrock Do(a/b), Hy, and Joa, Duffy Fy(a/b), Fy(x) and GATA Silencing, Kell Js(a/b), K/k, and Kp(a/b), Kidd Jk(a/b), Knope Kn(a/b), Landsteiner-Wiener LW(a/b), Lutheran Lu(a/b), MNS MN (GYPA n59, GYPA n71/72), and MNS Ss and U (Intron5, n1143, n230), Rh RHCE C/c, E/e, L245V, M238V, G336C, Scianna Sc(a/b), Yt(a/b). Challenges overcome with this panel are extensive homologies in the MNS group between Glycophorin A, B, and E genes and in the Rh group between the RHCE and RHD genes that pose challenges to assay development. The OpenArray plates are run on the QuantStudio™ 12K Flex with a throughput of roughly 400 samples in 3.5 hours. Genotypes were successfully validated on control samples. Blood Group Genotyping on the OpenArray platform provides an accurate tool for high-throughput screening of blood units which can be a complementary approach to existing serology tests.

Clinical Validation of Noninvasive Prenatal Screening for Fetal Sex Chromosome Aneuploidies in Maternal Plasma using Direct Analysis of Selected Regions (DANSR™) Assays.


Introduction: Fetal cell-free DNA (cfDNA) in maternal plasma enables techniques for screening for fetal aneuploidy using next generation sequencing technologies. We have previously described a method using DANSR assays for biochemical analysis of chromosomes 13, 18, and 21, combined with the Fetal fraction Optimized Risk for Trisomy Evaluation (FORTE™) algorithm to compute the risk of trisomy based on the DANSR assay results with high sensitivity and specificity. Both components are integral parts of the commercially-available Ariosa Harmony Prenatal™ Test. In this study, we have developed additional DANSR assays for the X and Y chromosomes and have applied the FORTE algorithm on a blinded set of samples. The set contained samples with and without sex chromosome aneuploidies (SCA).

Objective: Report clinical validation results on Harmony Prenatal Test’s ability to detect fetal SCA. Methods: 432 subjects were selected. Participants provided informed consent under an IRB approved protocol. All subjects underwent invasive testing. Resulting karyotypes were used to confirm the Harmony Prenatal Test results. Samples were processed as previously described with lab and analysis personnel blinded to the fetal karyotype. FORTE models were built against monosomy X, XXX, XXY, and XXXY genotypes. Results were compared against the fetal karyotypes. Results: Of the 414 plasma samples that passed standard QC metrics for the Harmony Prenatal Test, all generated a sex chromosome result (100%; 95% CI: 99.1-100%). All samples were concordant with the fetal karyotyping result for fetal sex (100%; 95% CI: 99.1-100%). 26 of 27 monosomy X samples classified as high risk agreed with karyotype (sensitivity 96.3%; 95% CI: 81.7-99.8%), with two discordant high risk results (specificity 99.5%; 95% CI: 98.1-99.9%). One XXX sample classified as high risk was concordant with karyotype with two discordant high risk results (specificity 99.5%; 95% CI: 98.1-99.9%). All of six XXXY calls were concordant (sensitivity 100%; 95% CI: 61-100%; specificity 100%; 95% CI: 99.0-100%). Conclusions: FORTE analysis of cfDNA with DANSR assays allows for risk assessment of nonmosaic fetal SCA. While this is the largest fetal SCA validation study done to date, due to the complex nature of SCAs, and possible maternal mosaicism, future larger studies are warranted. This study also demonstrates the ability to expand the Harmony Prenatal Test to genetic conditions other than trisomies 13, 18 and 21.

Next Generation ABO Genomics: the NHLBI Exome Sequencing Project (ESP).

K. Fox1, I. Stanaway2, M.B. Bamshad3,4, P.L. Auer5,6, A. Gordon7, D. Crosslin1, M. Foranga1, D. Green1, S. Rich1, A.P. Reiner1,3, D.A. Nickerson1,2, J. Johnsen7,8,10, 1) University of Washington, School of Medicine, Department of Genome Sciences, Seattle, WA, USA; 2) University of Washington, School of Medicine, Department of Pediatrics, Seattle, WA, USA; 3) Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA, USA; 4) University of Wisconsin-Milwaukee, Zilber School of Public Health, Milwaukee, WI, USA; 5) The University of Texas Health Science Center at Houston, Institute of Molecular Medicine, Research Center for Human Genetics, Houston, TX, USA; 6) Northwestern University, Feinberg School of Medicine, Chicago, IL, USA; 7) University of Virginia, School of Medicine, Center for Public Health Genomics, Charlottesville, VA, USA; 8) University of Washington, School of Public Health, Epidemiology, Seattle, WA, USA; 9) University of Washington, School of Medicine, Hematology Division, Seattle, WA, USA; 10) Puget Sound Blood Center, Research Institute, Seattle, WA, USA.

Cardiovascular Disease (CVD) is the leading cause of death in the United States. Recent studies demonstrate that additional ABO variants (e.g., non-ABO blood groups) can determine the risk for CVD. ABO phenotypes are the result of genetic variation in the coding portion of the ABO gene on 9q34.2. The influence of ABO subtypes (e.g., the A1 and A2 haplotypes) on CVD risk has yet to be fully explored. Previous studies have genotyped SNPs that tag ABO rather than defining the haplotype sequence necessary for subtype detection. We have developed an accredited ranking method to assign ABO haplotype using exome sequence data. This new method uses a matrix-based scoring approach to define subtypes from exome sequencing data to reference A, B, and O haplotypes from the Blood Gene Mutation Data Base (BGMut). We applied our method to phased exome data from ~5,600 individuals of European and African ancestry derived from the ESP. We observed 100% concordance between predicted and actual phenotype in those ESP participants with serological phenotyping data (n=680). We identified 24 common variants known to influence the function, including a common exon 6 indel that leads to the O phenotype, and another common indel that results in the A2 subtype. We have also identified rare coding variants within ABO (single nucleotide/misense variants, n=18), insertion/deletion/structiral variants (n=1), and structural variants (n=1) that segregate on known haplotype blocks. This method has the potential to improve the specificity of blood typing at both the clinical and research level and to reveal novel associations between CVD and previously unidentified rare ABO haplotypes.
2606W

Genome Analysis of iPS Cells for Regenerative Medicine. A. Watanabe1, N. Amano1, M. Nakamura1, A. Fukuhara1, P. Uynane1, Y. Tokunaga1,2, M. Yamaguchi1, T. Aoi1, K. Okita1, K. Takahashi1, S. Yamanaka1. 1) Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan; 2) Amelieff.

Pluripotent stem cells are now suggested as an artificial source of tissues, and consequently it is necessary to be able to guarantee their safety in the human body after transplantation. However, both embryonic stem (ES) cells and induced pluripotent stem (iPS) cells are produced after long-term culture, and thus harbor clone-to-clone variations in their DNA sequences and copy numbers as well as epigenetic profiles. It is therefore important to validate the quality of ES and iPS cells by genomic analyses. We aim to establish the standard method for evaluating iPS cells as a clinical-grade cell source by genome and epigenome analysis. We built the pipeline for both single-nucleotide variations (SNVs) and DNA copy number variations (CNVs), and performed exome and whole genome resequencing and compared between the original somatic cells and established iPS cells. We found not only clone-to-clone variations in iPS cell clones but also the heterogeneity in original somatic cells, indicating that comparison between original cells and established cells is essential for evaluating pluripotent stem cells. One of the iPS clones shows no non-synonymous mutations in exonic region. We also performed DNA methylation analysis by sequence-capture-based deep sequencing and identified novel DNA methylation sites specific for iPS cells. We introduce the strategy of genome and epigenome analysis for evaluating iPS cells for regenerative medicine using iPS cells.

2607T

Utility of targeted inherited disease panels for the diagnosis of rare congenital, potentially genetic syndromic disease. A. Khromykh1,2, D. Thach1,2, M. George1,2, J. McCartney1,2, W. Wong1,2, R. Baveja3,4, R. Iyer1,2,1) Inova Translational Medicine Institute, Inova Health System, 3300 Gallows Road, Falls Church, VA 22042; 2) Inova Fairfax Hospital, Inova Health System, 3300 Gallows Road, Falls Church, VA 22042; 3) Inova Fairfax Children’s Hospital, Inova Health System, 3300 Gallows Road, Falls Church, VA 22042; 4) Fairfax Neonatal Associates, 2730-B Prosperity Avenue, Fairfax, VA 22031.

Specific and accurate diagnostic tests are available for almost 3000 genetic disorders. Nevertheless, for many children with rare, potentially genetic syndromic disease, clear diagnosis remains elusive, and the diagnostic odyssey continues. Recently, next generation sequencing has been used to discover genes involved in rare Mendelian disease and for their diagnosis. However, there are significant barriers for equal and widespread accessibility to these tests, including high cost of testing and reimbursement issues, lack of clearly defined, validated, available algorithms to identify clinically relevant results and the need for multidisciplinary expertise to interpret the results. A short term and partial solution to some of these issues may be to implement a tiered testing modality. This includes 1st tier testing on commercially available, cheaper, targeted disease-gene panels which provide less complex and easier to interpret results using algorithms provided by the manufacturer or developed in-house. Though such testing will not identify new genes for Mendelian disease, it could help in the elucidation of new phenotypes associated with targeted genes present in the panel, reduction of further complex and expensive targeted testing, and elimination of the diagnostic odyssey for some families. In this study, we are analyzing our expanding cohort (currently 40+) of families with a proband afflicted with severe multiple congenital malformations that are not consistent with any known clinical diagnosis using targeted gene panels. The probands are negative by currently available standard genetic testing including arrays (when indicated and performed); most of these have not been analyzed on clinically available next generation panels, and have no clearly relevant family history. Targeted sequencing of the coding exons of >800 mendelian disease genes associated with rare and ultra-rare neuromuscular, cardiovascular, developmental, metabolic and other inherited disease, is performed using the Ion Amplicon and TruSight inherited disease panels (by Life Technologies and Illumina respectively). We expect that our study will provide useful information regarding the analytical and clinical sensitivity, specificity and utility of these panels that could provide support for the deployment of such tiered testing strategies in the clinical laboratory, using these and/or other enhanced targeted gene panels.

2608F

Towards a medical grade exome: Use of a gold standard to evaluate and enhance exome sequencing for diagnosis. M. Pratt1, G. Bartha2, S. Lu1, M. Harris1, S. Garcia1, G. Chandrakilake1, S. Cheruit2, R. Chen2,4, M. Clark1, M. Snyder2,3, J. West1, R. Chen1. 1) Personalis Inc, Menlo Park, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Stanford Center for Genomics and Personalized Medicine, Stanford University, Stanford, CA; 4) Icahn School of Medicine, Mount Sinai, New York, NY.

Exome sequencing is increasingly utilized in clinical genetics practice to diagnose cases where other genetic testing has been unsuccessful or would be cost-inefficient. However, diagnostic yield estimates from clinical exome testing remain low (~25%). Furthermore, standards for exome sequencing have not converged with respect to content inclusion and minimum coverage requirements to achieve clinical sensitivity and specificity on that content. We have assembled a broad content set over which we assess and optimize exome performance to significantly improve coverage of medically interpretable content with the aim of increasing diagnostic yield. To assess absolute performance of exome sequencing approaches, we have developed internal gold standards based on genomic sequencing. As an accuracy standard, we use a reference comprising more than 3 Tb of next generation sequencing data over multiple technologies on a large pedigree to develop a thorough set of confirmed inherited variants comprising single nucleotide, indel, and structural variants. We also compute coverage standards across samples at nucleotide and exon resolution over whole genomes and exomes. Using these standards, we determined sensitivity and specificity over the interpretable exome at varying mean sequencing depths in order to develop coverage targets by region and content type to ensure analytic validity over the most interpretable regions. We further assessed performance on additional functional and non-functional, e.g. UTRs, regions. Through examining coverage and accuracy shortfalls over interpretable content, we developed additional pull-down targets and protocols to augment commercially exomes in regions of low (<20X) or absent coverage with the aim of creating a finished medical-grade exome. We tested regulatory regions and other non-exonic regions known to contain disease and pharmaco-genetic-associated variation. This approach greatly increases the number of biologically-relevant genes finished at our medical coverage target (approximately >200x). Increased coverage enables sensitive and specific variant detection including at an additional 4854 loci having previously reported deleterious variants. Assessing accuracy on a larger set of biomedical variants including regulatory region variants and those associated with complex disorders, we increase sensitivity from 58% to 90% while reducing the error rate.

2609W

Genetic Research and diagnostic using Fluidigm Integrated Fluidic Circuits (IFCs), D. Bercovich1, Y. Plostky1, S. Alon-Shalov1. 1) Human Molecular Genetics & Pharmacogenetics, Migal- Tel Hai, Kiryat Shmona, Galilee, Israel; 2) Galil Genetic Analysis - GGA lab, Kazarin, Israel; 3) The Institute for Genetics, Ha’Emek Medical Center, Afula 18101, Israel.

The flexibility of the BioMark Real-Time PCR System, allow us to perform genetic research using different types of nano-fluidic (48.48 or 96.96) chips setup, in the thermal cycle of these chips and image the data in real time for quantity determination of DNA copy numbers (CNV) or mRNA expression in multiple loci locations or genes, and can also be used as an endpoint image reader for analyzing different allelic genotyping frequencies in a panel of 90 common mutations, in the Israeli populations, for 42 different mono genetic disorders revile there frequency of 15% in screening over 250 individuals general Israeli population , and 15-14 common mutations in Breast/Ovarian and Colon cancers revile there frequency of 3-4% in the Israeli population. A panel of 47 novel Canis (dogs) SNPs was used, for the determination of a phylogenetic tree in 45 different Canis DNA samples. Digital-PCR gene sets were used to determine the number of a human gene which were incorporated in a plant cells for the production of this protein to treat human disease.
2610T Whole Exome Sequencing is a sensitive and cost-effective method of detecting mutations in Osteogenesis Imperfecta and Marfan syndrome. A.M. McInerney-Leo 1, M. Marshall 1, B. Gardiner 1, P. Coucke 2, B. Loesy 2, J. West 3, M. West 3, B.P. Wordsworth 4, A. Zanki 5, P.J. Leo 1, M.A. Brown 1, E.L. Duncan 1. 1) University of Queensland Diamantina Institute, Wooloongabba, Queensland, Australia; 2) Department of Medical Genetics, MBF, 1st floor Ghent University Hospital De Pintelaan 185, B-9000 Ghent, Belgium; 3) School of Medicine, The University of Queensland Coordinator, Cardiovascular Genetic Clinic, The Prince Charles Hospital, Rode Road, Chermside QLD, Australia; 4) Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford England; 5) Bone Dysplasia Research Group, UQ Centre for Clinical Research (UQCCR), The University of Queensland, Building 71/918, Level 3 Royal Brisbane and Women's Hospital, Herston, QLD Australia.

Osteogenesis imperfecta (OI) and Marfan syndrome (MFS) are amongst the commonest Mendelian disorders. Both are normally diagnosed clinically without genetic testing, as conventional sequencing is expensive due to the size and number of potentially causative genes and mutations. However, genetic testing benefits patients, at-risk family members and individuals with borderline phenotypes; improves genetic counselling; and allows critical differential diagnoses. We assessed whether whole exome sequencing (WES) is a sensitive method for mutation detection in OI and MFS. Methods: WES was performed on genomic DNA from 13 participants with OI and 10 participants with MFS, all of whom had known mutations, with massive parallel sequencing of multiplexed samples. SNPs and small indels were called using Genome Analysis Toolkit (GATK) and annotated using ANNOVAR. CREST, exomeCopy and ExomeDepth were used to detect larger deletions. Results were compared with the previous data. The target capture of the currently available exome capture platforms was also compared. Results: All 13 mutations in the OI cohort and 9/10 in the MFS cohort were detected (overall sensitivity >95.6%) which included 11 previously unreported mutations, including 2 deletions. One mutation was not detected by GATK due to strand bias. Capture platforms and analysis programs differed considerably in their ability to detect mutations. Costs for WES and the available commercial screening programme are comparable. Conclusion: WES is both sensitive and cost-effective for mutation detection in patients with OI and MFS. Careful selection of platform and analysis programs is necessary to maximise success.

2612W High-throughput molecular genetic analysis in 92 patients with steroid resistant nephrotic syndrome applying Fluidigm Access Array™ technology. A. Alissa 1,2, A. Tahir 1,2, J. West 3, M. Marshall 1, B. Gardiner 1, P. Coucke 2, B. Loesy 2, J. West 3, M. West 3, B.P. Wordsworth 4, A. Zanki 5, P.J. Leo 1, M.A. Brown 1, E.L. Duncan 1. 1) University of Queensland Diamantina Institute, Wooloongabba, Queensland, Australia; 2) Department of Medical Genetics, MBF, 1st floor Ghent University Hospital De Pintelaan 185, B-9000 Ghent, Belgium; 3) School of Medicine, The University of Queensland Coordinator, Cardiovascular Genetic Clinic, The Prince Charles Hospital, Rode Road, Chermside QLD, Australia; 4) Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford England; 5) Bone Dysplasia Research Group, UQ Centre for Clinical Research (UQCCR), The University of Queensland, Building 71/918, Level 3 Royal Brisbane and Women's Hospital, Herston, QLD Australia.

Steroid resistant nephrotic syndrome (SRNS) is a genetically heterogeneous kidney disease characterized by heavy proteinuria, hypoalbuminemia, and edema. In most SRNS cases the kidney function declines over time, resulting in end-stage renal disease which necessitates renal replacement therapy. Mutations in more than 20 genes are implicated in the pathogenesis of SRNS rendering mutational analysis tedious and costly when applying conventional Sanger sequencing. Here, we screened for mutations in 21 established SRNS genes in 92 mostly pediatric patients recruited at the University of Michigan with SRNS. Methods: We implemented the PCR-based Fluidigm Access Array™ approach established previously in our lab 1. We screened for mutations in all coding regions (424 exons) of 21 established monogenic SRNS genes (NPHS1, NPHS2, PLCE1, COQ2, COQ6, LAMB2, WT1, PDSS2, CD2AP, SMARCA1, PTPRO, CFH, CUBN, SCARB2, INF2, MYO1E, NEIL1, TRPC6, ACTN4, LMIXB, and ITGA5) in a cohort of 92 mostly sporadic childhood cases with SRNS. We applied the Fluidigm Access Array™ nanofluidic technology followed by barcoding and next-generation sequencing on an Illumina HiSeq2000 platform. Bioinformatics analysis was performed using CLC Genomics Workbench™ software and significant variants were validated by Sanger sequencing. Results: We established the molecular diagnosis in 12 out of 92 patients (13%) and detected pathogenic mutations in the genes NPHS1, NPHS2, WT1, PLCE1, LAMB2, and ACTN4. Moreover, we discovered novel mutations in the genes NPHS1 (p.Tyr638Cys, p.Gly601Ala), WT1 (p.Tyr927X), PLCE1 (p.Ser1190Gly), LAMB2 (p.Ile992Thr), and ACTN4 (p.Arg656Gln). Conclusion: Fluidigm Access Array™ high-throughput mutation analysis allows screening of a large patient cohort for multiple genes in parallel at low cost and helps to find mutations in the undiagnosed patients in SRNS. Halbritter et al., Hum Genet 2013, Apr 5 [Epub ahead of print].

2611F Rapid and high mutation detection rate using ion torrent technology and inherited disease panel. N. Al Tassan 1,2, A. Almstafa 1,2, D. Khalil 1,2, J. Shinwari 1,2, R. Kattan 1,2, A. Alissa 1,2, A. Tahiri 1,2, M. Abouelhoda 1,2. 1) Department of Genetics, King Faisal Specialist Hospital, Riyadh, 11211, Saudi Arabia; 2) Saudi Genome Project, King Abdulaziz for Science and Technology, P.O Box 6086, Riyadh 11442, Saudi Arabia.

Ion Torrent semiconductor™ is a chip based sequencing technology (Life technologies, Guilford, CT) that performs sequencing-by-synthesis accompanied by electrochemical detection of base incorporation. In addition to Whole genome and exome sequencing, the technology is being applied in a heavily multiplexed PCR-based approach (Ampliseq) for targeted gene panels. The Ion Ampliseq™ Inherited Disease Panel (IDP) uses >10,000 primer pairs divided into 3 PCR pools to amplify the exons of 328 genes associated with approximately 700 inherited diseases. We applied Ion Torrent sequencing and this gene panel to study 100 samples having a Sanger sequencing validated mutation in one of the genes covered by the panel. 10ng of each sample was used to the Ampliseq PCR, the 3 PCR products were pooled and the library was purified, amplified and quantified. 20 pmoles of each library was used for emulsion PCR, followed by enrichment and sequencing on the medium scale 316 Chip for single sample or the high scale 318Chip for multiplexed samples. Specificity and sensitivity was 100% for all homoyzgous single point mutations. Several indels required manual review of to identify the mutation which was not identified using the current variant caller. Given consanguinity of the study population and consequent frequency of homoallelic mutations, we also examined exclusion of genes based upon presence of heteroallelic SNPs. Ion Torrent technology and sequencing of highly multiplexed panels offers an efficient screening/diagnostic option for inherited diseases commonplace in consanguineous populations.
2614T

Mutaome Profiling and Retrospective Mutaome Profiling Using Archived Bone Marrow or Peripheral Blood Smear in B-ALL. F. Wang, H. Liu, W. Teng, Y. Wang, X. Chen, L. Guo, M. Wang, Q. Yin, H. Yang, F. Zhu. 1) Molecular Medicine Lab., Hebei Yanda Hospital, Sanhe, Hebei, China; 2) Department of Hematology, Peking University First Hospital, Beijing, China.

Background: B-cell acute lymphoblastic leukemia (B-ALL) is a heterogeneous disease with respect to presentation and clinical outcome. In recent years, more and more somatic mutations and their clinical significance were identified in B-ALL, including point mutations and large genomic sequencing deletion mutations. Herein we recommend using the novel word "mutaome" for representing the repertoire of somatic gene mutations in a specific tumor tissue, and aimed to establish a panel of mutation profiling protocol for clinical use in B-ALL. Cases and Methods: Bone Marrow (BM) or peripheral blood (PB) was collected form patients with newly diagnosed or relapsed. Archived BM or PB smear on glass slides at the time of newly diagnosed were used for retrospective mutation analysis. Mutation profiling protocol for JAK1, JAK2 (Exon16, 20), PA5X, PHF6 and TP53 by PCR and Sanger sequencing was established, with the detection sensitivity about 15% to 20%. IKZF1 exon deletion, PA5X exon deletion and MLL partial tandem duplication (MLL-PTD) were detected by RT-PCR, electrophoresis and sequencing of the aberrant bands, which didn't applied to archived smear specimens. Results: 1) Totally 24 archived smear, 29 fresh BM or PB samples from 53 patients were analyzed. Age ranged from 2 to 81 years old, with the median age of 14 years old. 2) Totally 15 out of 29 fresh BM or PB samples carrying one or more mutations. The number of mutated cases for each gene is 5 for TP53, 3 PHF6, 1 FLT3, 3 PA5X exon deletions and 4 IKZF1 exon deletions. Totally 6 out of 24 archived smear specimens carrying one or more mutation, 2 with TP53 mutations, 3 with PA5X point mutations, 1 with FLT3 and 1 with PHF6 mutations. Conclusions: Panel testing for mutations is effective method for detection of B-ALL molecular markers. Archived BM or PB smear samples can be used for retrospective mutaome analysis.

2615W

Molecular characterization of children with severe autism spectrum disorders. A.C. Tsai. OHSU, Oregon, OR.

From April 2012 to May 2013, 10 pts with severe autism spectrum disorder ascertained in the autism and general genetic clinic in the CDCR/DCH were characterized by CMA and molecular panels, the clinical features of the patients in his cohort fall in the Angelman/Rett syndrome-like phenotype: patient either present with severe seizure at birth, severe hypotonic, being non-verbal after 7 years of age, some with history of sleep apnea, unique hand mannerism, ataxia, fascination about the water or inappropriate laughter. Method: All patients receive Rett/Angelman-like panel followed by microarray if negative. The panel includes CDKL5, SLC9A6, TCF4 and FOXP1, UBE3A and MEC2P. When the results of the above tests are normal, additional testing based on clinical finding were added including FLNA, RAI1 and seizure panels, as well as SNP plus 400K CMA. Results: 7 Patient were found to have remarkable findings which include c.409T>C heterozygousmutation of UBE3A, MEC2P (Zappella variant), duplication of MEC2P, FOXP1 1-bp deletion, PQBP1 1-bp insertion, and 2 with 1p36 deletion, 6 with 1p36 deletion, 13 with 1q21.3 deletion of 14q13.2 with intragenic deletion of the RALGAPA (aka TULP1; aka GARNL1) Discussion: the significance of c.409T>C heterozygous mutation of UBE3A is still unknown; this girl is non-verbal but has significant signing, the rest of the testing were normal MOCA carries this mutation but cannot determine the parental origin as maternal grandfather is not available for testing. The TULP1 deletion is inherited from the father who has autistic features but clinically is less involved than the proband. This child also received a seizure panel which were negative. The child with PQBP1 mutation was originally thought to have angelman syndrome but detected by the seizure panel which include the AS/Rett panel, the diagnosis was subsequently changed. Conclusion:This small study highlight the clinical utility of panel assay which is perfect to characterize conditions with a spectrum of presentation in a gene and overlapping phenotype among the genes on the pathway. This study also demonstrate the importance of proper pre-testing genetic counseling, as the test result might detect unexpected genetic syndromes.

2616T

Establishment of a next-generation sequencing protocol for genetic testing of tuberous sclerosis complex. P. Chen1,2,3,4, 1) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 2) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan.

Tuberous sclerosis complex (TSC) is a characteristic disease drawing high clinical and research attention. It is an autosomal dominant disorder characterized by tumors in different body organs, including brain, kidney, liver, skin, lung, heart and eye. TSC is caused by dysfunction of the critical biological pathways related to mammalian target of rapamycin (mTOR). About 400 different causative variants have been identified in the two known genes (TSC1 at chromosome 9q34 and TSC2 at chromosome 16p13). However, routine genetic testing for TSC patients using traditional Sanger sequencing method is too expensive and labor intensive because there are a total of 64 exons of these 2 genes. We decided to apply the NGS technology for the cost-effective diagnosis of TSC. We have focused on the whole genomic regions containing all exons, all introns and 10 Kb upstream and 10 Kb downstream of the TSC1 (hg19, chr9:135761735-135830020, 68285 bp) and TSC2 (hg19, chr16:2087990-2143712, 55722 bp) genes. A Roche NimbleGen customized capture library was designed for enrichment of the targeted sequences. A reasonably good coverage (97% of the intended regions; and 100% of the exons) can be achieved. The enriched libraries were then pair-end sequenced on the Illumina HiSeq2000 Genome Analyzer. The whole bioinformatics analysis was processed in a computer cluster with more than 50 nodes at the High Performance Computing (HPC) center at National Taiwan University. We used BWA for initial read mapping, GATK pipeline for realignment and SNP/indel calling. Aggressive filtering was applied to further reduce false positive results. Biological significance of the genetic variants was then predicted using bioinformatics tools, mostly SIFT and PolyPhen2. The integrative genomics viewer (IGV) was used for visualization of the results. Among the 61 TSC patients we tested, 50 (82%) of them could be assigned causative variants. We demonstrated that our new protocol can detect not only the single nucleotide substitution and small indels but also difficult genetic variants (such as a big deletion of ~15 Kb size, and mosaicism). This serves as a single stop for TSC genotyping and waive the necessity of both Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA). We consider this new protocol a cost-effective diagnostic test for TSC, and may be applied to other diseases in the future.
5, 7
and equipoise in both primary care and cardiology clinics within the MedSeq Project. We describe the evaluation of the effects of CRR disclosure on additional testing and discuss the role of small effect-size risk in clinical decision making. We will present a summary of the quantitative risk, even from a limited number of small effect loci, was comparable in clinical trial support. For the eight phenotypes in the current CRR, (abdominal aortic aneurysm, atrial fibrillation, coronary heart disease, type 2 diabetes, hypertensive disease, atrial septal defect, and primary open-angle glaucoma (POAG)).

Genome-wide association studies have defined the common risk alleles for numerous human traits, but to date these findings have not impacted clinical practice. The MedSeq Project is a randomized clinical trial that aims to develop standards and procedures for the evaluation and reporting of genome sequencing data and will directly assess the impact of integrating genome sequencing into clinical medicine. MedSeq participants in the study will receive both a General Genome Report and a Cardio Risk Report (CRR), the latter featuring a summary of variants in a targeted panel of 102 monogenic cardiac disease-associated genes, a polygenic predicted fasting lipid profile, and common allele risk information for eight common cardiovascular phenotypes. To define a rational basis for including common alleles in the CRR, we estimated the effect sizes for rigorously validated common risk alleles across all published loci for all phenotypes directly or indirectly related to cardiovascular disease and compared these with the estimated effect sizes of non-genetic risk factors commonly used in clinical decision-making, but which have no randomized clinical trial support. For the eight phenotypes in the current CRR, (abdominal aortic aneurysm, atrial fibrillation, coronary heart disease, type 2 diabetes, hypertensive disease, atrial septal defect, and primary open-angle glaucoma (POAG)),

ClinVar: Improving Access to Clinically Relevant Variants for the Research and Clinical Genomics Communities. M.J. Landrum, J. Lee, G. Riley, R. Tully, S. Chitipiralla, M. Halavi, D. Hoffman, J.B. Holmes, W. Jang, K. Katz, M. Ovetsky, A. Sethi, R. Villamarin, D.M. Church, W.S. Rubinstein, D.R. Maglott. ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) maintains a freely available, public repository for the relationships between variants and phenotypes along with supporting evidence. The database can be used interactively or incorporated into variant analysis pipelines. The public release of the database in April 2013 largely included curated data from OMIM and GeneReviews, as well as submissions from some LSDBs, testing labs, and ClinSeq. Within the first two months, submitters also included additional testing labs, research groups, and expert panels and professional societies.

Advances in technology for genomic testing, particularly whole exome and genome sequencing, are identifying thousands of new variants. However, determining the clinical relevance of variants can be challenging, especially for rare variants. To increase the accuracy of such data, ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) provides a comprehensive approach to reducing glaucoma blindness by unraveling the genetic complexity of glaucoma, providing accredited genetic testing and genetic counseling, and translating research findings in better clinical care.


Purpose: Glaucoma is the leading cause of irreversible blindness worldwide and is a complex disorder with genetics playing a crucial role. Adequate monitoring and interventions at the early stages can prevent glaucoma blindness. Genetic testing is promising strategy to identify at risk individuals and reduce the impact of glaucoma blindness through prevention. The Australian and New Zealand Registry of Advanced Glaucoma (ANZRAG) has established a biobank of severe glaucoma cases to identify novel genetic risk factors for the worst glaucoma outcomes, and to establish genetic testing protocols for known glaucoma genes. Methods: Advanced-Open-Angle Glaucoma (OAG) cases defined by central visual field loss or severe peripheral vision loss were recruited. Secondary glaucoma cases were also recruited regardless of severity. Cases with advanced glaucoma and Primary Congenital Glaucoma (PCG) were tested in an accredited pathology laboratory by direct DNA sequencing for known glaucoma genes, Myocillin and CYP1B1 respectively, and results were provided back to the participants with the provision of genetic counseling. Cascade genetic testing was made available for relatives of participants with pathogenic mutations in these genes when clinically relevant. Results: 1570 participants with severe OAG and 960 with other glaucoma subtypes have been recruited. Using genome-wide association studies, new glaucoma risk alleles were identified: TMCO1 in advanced primary OAG cases, and CDKN2B-AS1 associated with normal-tension glaucoma. Myocillin mutations were identified in 4.2% of advanced primary OAG individuals with 1.6% sharing a paternal inheritance of the OAG. These mutations were found in individuals with younger age at diagnosis and high-tension glaucoma. CYP1B1 mutations were found in 18% of PCG cases. Cascade genetic testing has identified highly penetrant Myocillin mutations in 35 as yet asymptomatic individuals. The experience of tested asymptomatic individuals has been evaluated to help health professionals in providing better support to patients. Outcome: The ANZRAG provides a comprehensive approach to reducing glaucoma blindness by unraveling the genetic complexity of glaucoma, providing accredited genetic testing and genetic counseling, and translating research findings in better clinical care.
2620F Molecular diagnostic approach for limb-girdle muscular dystrophy using both multi-gene panel sequencing and Sanger sequencing. H. Park1, S. Lee1, S.H. Seo1, S. Park1, S.I. Cho2, M.W. Seong1, S.S. Park1 1) Department of Laboratory Medicine, College of Medicine, Seoul National University Hospital, Seoul, Korea; 2) Department of Laboratory Medicine, College of Medicine, Konkuk University Medical center, Seoul, Korea.

Introduction: Limb-girdle muscular dystrophy is genetically heterogeneous disease, with clinical involvement typically limited to skeletal muscle. Because more than 20 genes identified, it is not easy to find the causative mutations by conventional PCR and direct sequencing for individual patient. Here we present a molecular diagnostic strategy combing multi-gene panel sequencing and Sanger sequencing. Methods: Seventeen patients who were suspected of limb-girdle muscular dystrophy were screened for sixteen genes using TrueSeq Custom Enrichment Kit (illumina) and MiSeq (illumina). All previously reported mutations and probable pathogenic variants including novel non-synonymous ones were confirmed by Sanger sequencing. Also all low-coverage regions with coverage depth under 10X were resequenced by Sanger sequencing. Results: Among 17 patients, five ones were molecularly confirmed as limb-girdle muscular dystrophy. Two patients had a LMNA gene mutation (c.1357C>T and c.1386A>C), which was inherited autosomal dominant manner. Another two patients had compound heterozygous mutations in DYSF gene (c.2248C>T; 5668-7G>A) and c.2494C>T; (c.400delC) and the other one patient had two mutations in CAPN3 gene (c.439C>T; c.1076C>T) which were inherited autosomal recessive manner. Conclusion: Our multi-gene panel detected pathogenic mutation(s) in five limb-girdle muscular dystrophy patients. Multi-gene panel sequencing using next generation sequencing technology can be a cost efficient and fast method for diagnosis of genetically heterogeneous disease like limb-girdle muscular dystrophy.

2621W Coding mutations and variations in the 3'UTR of CYP21A2 gene in heterogeneous females associated with hyperandrogenism. V. Neocleous1, C. Shammas5, AAP. Phedonas1, M. Policò2, TC. Kyriakides1, M. Tounba3, N. Skordis2, LA. Phylactou1. 1) Molecular Genetics, Function & Therapy, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 2) Allithias Endocrinology Center, Nicosia, Cyprus; 3) Department of Epidemiology & Public Health, Yale University, USA; 4) Iasis Hospital, Paphos, Cyprus; 5) Paediatric Endocrine Unit, Makarios III Hospital, Nicosia, Cyprus.

Congenital adrenal hyperplasia (CAH) is a common autosomal recessive disorder primarily caused by mutations in the CYP21A2 gene. Heterozygosity for CYP21A2 mutations in females increases their risk of clinically manifesting hyperandrogenism. The present study was designed to seek evidence on the association between the mutations in the CYP21A2 gene and the biochemical/clinical findings on heterozygous children, adolescents and women with hyperandrogenemia. Moreover, the implication of variants in the 3'UTR region of the CYP21A2 gene was investigated. The hormonal response to ACTH was evaluated in heterozygous females with clinical signs of hyperandrogenism along with direct DNA sequencing and Sanger sequencing. Methods: In total, 33 female patients were enrolled in the study. The CYP21A2 gene was sequenced using Ion AmpliseqTM PCR and then sequenced by PGM. Meanwhile, a set of mutant sample mixture representing common mutations in patients were prepared, and the PGM sequencing might be suitable for screening gene mutation in HPA and the severity of CAMT supports the value of including testing for this disorder in future AJ population screening.

2622T Detection of deleted D4Z4 locus in turkish patients with facioscapulohumeral muscular dystrophy. S. Berkier Karaaouz1, OB. Sahhan1, H. Yalcin1, OB. Sahan1, S. Birke 1) Med Biol& Gen, Akdeniz University Faculty Med, Antalya, Turkey; 2) Dept. of Neurology Akdeniz University Faculty Med, Antalya, Turkey.

Facioscapulohumeral Muscular Dystrophy (FSHD) which is characterized by progressive weakness of muscles in the face, shoulder girdle and upper arms is the third most common muscular dystrophy. FSHD is caused by deletions of the D4Z4 repeats in the 4q35 region. In this autosomal dominantly inherited disease, more than 95% patients have only 1-10 repeats instead of 11-100 repeats observed in healthy controls. In this study, 24 individuals (11 male, 13 female) from 6 unrelated Turkish families with FSHD were handled after the preclinical diagnosis by the Akdeniz university medical faculty, department of neurology. Three healthy individuals were studied as a control group. In order to show the deletion of D4Z4 tandem repeats at the q35 locus on chromosome 4, Southern blot method was performed. While D4Z4 repeats of control group was found to be in the normal range (more than 11 repeats), 20 FSHD patients’ repeat numbers were observed to be under the normal range (less than 11 repeats) as discordant to clinical findings. Our results show that, Southern blot was observed as a suitable method for determining the D4Z4 repeat deletions on chromosome 4 and also, because of in all the affected cases displayed D4Z4 repeat deletions, this analysis could be performed in patients with myopathies similar to FSHD, especially for prenatal diagnosis and genetic counseling.

2623F Investigating the carrier screening potential of the MPL c.79+2T>A transversion, a known cause of congenital amegakaryocytic thrombocytopenia for individuals of AJ descent. S. Birke1, J. Sugalski1, C. Holland1, J. Stoeker1, J. Buish, Ann Arbor, MI.

Mutations in the thrombopoetin receptor c-MPL cause a rare inherited disease known as congenital amegakaryocytic thrombocytopenia (CATM). This disease presents as severe thrombocytopenia at birth, reduced megakaryocytes and progression to bone marrow failure. The only known effective treatment for this a bone marrow transplant, which presents significant risk. Recently, a founder mutation in the c-MPL gene known as the c.79+2T>A transversion has been shown in the Ashkenazi Jewish population. Carrier screening for other disease alleles within this community has met with great success, reducing the numbers of new disease-affected individuals across a number of syndromes. Here, we present data from a blinded and random screening of individuals of AJ descent and compare this with a cohort of non-AJ individuals. In concordance with the previously studies, we find a significant carrier rate for the c.79+2T>A mutation amongst the AJ population. Our preliminary findings indicate that out of the first 358 tested, 6 are carriers for c.79+2T>A which have been confirmed via Sanger sequencing. We will plan on screening a total cohort of ~1000 individuals. The ability to screen for the c.79+2T>A mutation has been included in our validated AJP NXT carrier screen. Our data presented here along with previous studies and the severity of CATM supports the value of including testing for this mutation in future AJ population screening.

2624W Screening of gene mutation in hyperphenylalaninemia using Ion Torrent sequencing. Y. Cao, P. Song, Y. Qu, J. Bai, Y. Jin, H. Wang. Department of Laboratory Medicine, College of Medicine, Shanghai Jiaotong University School of Medicine, Shanghai, China.

Hyperphenylalaninemia (HPA) is a very common autosomal recessive genetic disease. It has been verified that the metabolic pathway of phenylalanine involves five pathogenic genes: PAH, GCH1, PTPS, QDPR, and PGM1. Rapid and accurate genetic diagnosis is very crucial for a clear diagnosis of disease types, choosing the right treatments in a timely manner, and for genetic counseling and prenatal diagnosis. To evaluate the feasibility of gene screening in children with HPA using Ion Torrent Personal Genome Machine (PGM), the CDS and UTR regions of PAH, GCH1, PTPS, QDPR, and PCBD1 in 3 patients with HPA and one healthy control were amplified using Ion AmpliseqTM PCR and then sequenced by PGM. Meanwhile, a set of mutant sample mixture representing common mutations in patients with HPA in China was used to evaluate the accuracy of the PGM. All of known mutation sites were correctly identified in the positive control. In addition, we detected 22 variations in the patients and the healthy control. Compared with database and verified by the Sanger sequencing method, it was confirmed that 6 were pathogenic mutations, 18 were polymorphisms and 4 were false-positive calls. Based on our study, the PGM sequencing might be suitable for screening gene mutation in HPA via metabolic pathways, which would meet the medical need for individualized diagnosis and treatment. However, mutations suspected for false-positive calls should be discriminated by other methods.
2625T
Comparison of Seven Commercial DNA Extraction Kits for the Isolation of Listeria monocytogenes DNA from Whole Blood Samples. M. El-Abdalla, T. A. Haj-Ahmad, Y. Haj-Ahmad, M. Al-Ati

1) Molecular Biology Department, National Research Center, Dokki, Cairo, Egypt; 2) Department of Biochemistry, Faculty of Science, Alexandria University, Egypt; 3) Department of Biomedical Sciences, University of Guelph, Guelph, Ontario, Canada; 4) Brock University, 500 Glenridge Avenue, St. Catharines, ON, L2S 3A1; 5) Norgen Biotek Corp., 3430 Schmon Parkway, Thorold, Ontario, Canada, L2V 4Y6.

Listeriosis is an important public health concern in North America; it is mainly transferred by food contaminated with the bacterium Listeria monocytogenes. The majority of susceptible individuals are pregnant women, newborns, older adults and those who are immunocompromised, with a mortality rate of about 30% among infected individuals. In this study we compared the recovery as well as the limit of detection of L. monocytogenes DNA from whole blood using seven commercial blood DNA isolation kits (QiaAmp DNA Blood Mini Kit, Norgen Genomic DNA Isolation Mini Kit, MoBio UltraClean DNA Blood Spin Kit, Macherey-Nagel NucleoSpin Blood Kit, Gentra Puregene Blood Kit, Norgen Non-enzymatic DNA Blood Kit and Norgen Enzymatic DNA Blood Kit). Human whole blood samples were spiked with known copy numbers of L. monocytogenes. Total genomic DNA was isolated and compared both qualitatively and quantitatively among the various kits; the concentration as well as OD260/280 and OD260/230 were measured spectrophotometrically. The isolated DNA was used in a real-time PCR reaction using specific primers. We determined the limit of detection for every kit as well as the linearity of pathogen recovery at different concentrations. Also we investigated the presence or absence of contamination and ease of handling of the used methods. Our results show that the column-based methods have higher consistency than alcohol precipitation methods, with higher yield and purity from procedures utilizing proteinase K.

2625W
A Case in Point: When is Extended Genotyping of AAT (SERPINA1) Indicated? S. Kwong, J. Stoller, F. Mularo, F. Lacbawan

1) Case Western Reserve University School of Medicine, Cleveland, OH; 2) Cleveland Clinic, Cleveland, OH.

Alpha-1 antitrypsin deficiency (AATD) (OMIM #613490) affects 1 in 2,000 to 1 in 5,000 individuals, but is clinically under-recognized. AATD is associated with mutations of SERPINA1, a serine proteinase inhibitor that functions as a major inhibitor of neutrophil elastase. AATD typically manifests with liver disease and/or early-onset emphysema. Patients who present with early-onset or unexplained (without typical risk factors for AATD) chronic liver disease, unexplained chronic obstructive pulmonary disease, and/or symptoms such as necrotizing panniculitis or vasculitis have alpha-1-antitrypsin levels measured. Abnormally low levels are followed up with S/Z genotyping or phenotyping with thin-layer isoelectrofocusing (IEF), the gold standard, to look for aberrant SERPINA1 alleles. The most well-understood and most common alleles are M (wild-type); S and Z ( dysfunctional mutant alleles). However, the Q0 (or null alleles), which have nonsense or frameshift mutations leading to absent or truncated mRNA transcripts or proteins that are incompletely characterized. Null alleles are typically not detected with conventional gold-standard methods and may require direct DNA sequencing.

Even then, the function of only a few null alleles has been well described, thus making predictions about their phenotypic manifestations difficult. The current report presents a case of an asymptomatic 48-year old female patient who had only chronically elevated AST and was found to have an AATM enzyme level of 54 mg/dL (normal range: 100-220 mg/dL). DNA sequencing showed heterozygous M1V (c.710 T>C) and S (c.863 A>T) alleles with heterozygous Q0_cairo (c.847A>T). Extended genotyping by subcloning fragment of SERPINA1 gene and transfection in HEK293T cells identified homozygous Q0_cairo. Since the compound heterozygote state produces the unexpected phenotype. To our knowledge, the Q0_cairo and S combination has never been described to date. Neither allele has been associated with liver manifestations, only with lung manifestations. A case demonstrating the importance of extended genotyping and the need to uncover rare alleles associated with AATD and to offer accurate genetic counseling information to the proband and family members. This newly described allelic combination extends the approximately 120 alleles described to date and permits better phenotypic characterization of individuals with unusual variants of AATD.

2626F

InVitae, San Francisco, CA.

Historically, diagnostic tests using DNA sequencing have only been offered for a limited set of genes to patients with specific clinical indications. The high cost of de facto standard assays (Sanger sequencing, MLPA, etc.), and more importantly, the high cost and challenges in clinical data interpretation have been cited among the reasons for this. Thus, many genetic diseases and clinically important genetic conditions often go undiagnosed. We have developed an in-house infrastructure for NGS-based diagnostic assay development, validation, and operation in a CLIA environment. To date, we have conducted a thorough scientific review of the literature for over 500 genes and their associated conditions, storing validated gene sequences, transcripts, risk models, and over 32,000 clinically characterized variants in a database used both to optimize assay design and to help interpret results. We have a hybrid calling and data QC pipeline employing GATK, Freebayes, and custom algorithms for different variant classes. Preliminary reports for known and novel SNVs, indels, and CNVs are automatically generated for review, and a team of medical specialists then classifies variants according to ACMG guidelines given the patient’s indication and signs-out finalized clinical reports. To date, validation has been performed on over 200 clinical samples. Get-RM and HapMap samples. 100% of clinical SNV and indel genotypes reported agreed with the results of established, traditional diagnostic assays on those samples. Importantly, pathogenicity assessments in the clinical reports agreed as well. In collaboration with other labs and patient advocates, we have launched an effort to expand the publicly available set of unpublished clinical variants that we believe will be critical in diagnostic settings. Most importantly, we believe these processes are highly unlikely to change the way the assay to grow to report on the vast majority of genetic conditions with high accuracy. For the HapMap samples specifically we performed a comparison against a combination of 1000 Genomes Project and Complete Genomics data. For all 2172 coding sequence SNVs in the 211 genes assayed in the HapMap samples, 99.7% sensitivity and greater than 99.99% specificity was demonstrated. For all 66 coding indels, 98.3% sensitivity and greater than 99.99% specificity was demonstrated.

2626G

1) National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD; 2) Office of Biotechnology Activities, National Institutes of Health.

The NIH Genetic Testing Registry (GTR; http://www.ncbi.nlm.nih.gov/gtr/ ) is a free, centralized, international registry of comprehensive genetic test information covering clinical and research tests for Mendelian disorders and drug response. As of June 2013, submitters have voluntarily provided detailed information on over 6400 registered tests for over 2700 conditions and 5600 genes. Active submitter participation coupled with GTR’s rich data structure enables many questions to be answered efficiently. Methods: Extensive stakeholder input and consultation with advisory boards determined concepts to be represented and ongoing feedback is elicited to enhance utility of GTR. All registered tests provide details about methodology, tested conditions, and relevant genes. Clinical tests have fields for purpose, test performance characteristics, target population, ordering information, AMA molecular pathology CPT codes, proficiency testing, and regulatory information. Research tests provide study description, eligibility criteria, consent form, and enrollment information. GTR staff augment content with practice guidelines, position statements and recommendations from expert sources and the website integrates information from authoritative sources. Results: As of June 2013, a total of 352 laboratories have registered 6465 tests employing molecular (N=6146; 95%), cytogenetic (N=154; 2.4%) and biochemical (N=287; 4.4%) methods (combinations included). Next-generation sequencing is a component of 6.0% of molecular tests. Eighteen registered labs offer whole genome or whole exome sequencing services. All clinical tests have analytical validity statements. Submitted data include statements with supporting citations [in brackets] about clinical validity (for 866 tests [281]), target population (for 1005 tests [305]) and clinical utility (632 total statements [329]). Among tests registered by US labs, 0.2% report FDA status cleared/approved. A total of 720 test performance guidelines, position statements, and recommendations pertaining to 289 conditions have been assembled by GTR staff. These results will be updated at the time of presentation. Conclusions: GTR enables informed selection of tests by clinicians, provides a snapshot of genetic testing, and helps identify knowledge gaps of interest to professional societies and regulatory agencies.

Clinical diagnostic exome sequencing (DES) provides data on all the coding exons of the genome and is currently indicated when prior tests have been negative, when the phenotype fits with the clinical spectrum of multiple genes, and/or when the phenotype is not consistent with any known clinical picture. DES has not only identified pathogenic alterations in genes not previously associated with the disease, but it has also revealed broader phenotypes not previously considered as part of the clinical spectrum within genes with well-established disease-associations. Even in cases where the patient has been provided with a clinical diagnosis or appropriate differential diagnosis, DES may still be pursued to identify the underlying molecular etiology when a single gene test was not available or when DES is more cost- and time-effective than the sequential gene-by-gene approach. A retrospective analysis of the first 200 reported patients undergoing DES at one laboratory revealed that prior to testing, 10% had a clinical diagnosis. Among the 20 clinical diagnoses made prior to DES testing, 3 (15%) were associated with 1-3 genes, while multiple genes could be implicated in the rest (85%). DES provided a definitive molecular diagnosis (including characterized and novel genes) in 53% of patients with a prior clinical diagnosis. The detection rate was lower (33%) in cases those with only 1-3 genes associated with the diagnosis, than those with multiple suspected genes (75%), thought to be attributed to the low overall clinical detection rate of the suspected gene(s). Gene coverage was greater than 95% for all the cases associated with 1-3 genes. These results highlight the clinical utility of DES, even among patients with a prior diagnosis, as it may reveal the underlying molecular etiology of the disease of interest when multiple genes may be involved.


The launch of Ambry’s cancer NG5 panels in early 2012 has led to unexpected genotype-phenotype discrepancies among patient results. We reviewed 38 ColoNext NG5 cancer panels that had mutations detected in genes associated with cancer syndromes with established clinical criteria. Of those, a significant percentage of patients with pathogenic mutations did not meet the correlating clinical criteria based on their personal and family history: 50% (3/6) with biallelic MUTYH mutations, 50% (2/4) with SMAD4 mutations, 50% (1/2) with PTEN mutations, 14% (1/7) with APC mutations, 40% (2/5) with MSH6 mutations, 28.5% (2/7) with MSH2 mutations, and 57% (4/7) with PMS2 mutations. Here we discuss specifics in twelve of these cases. Biallelic MUTYH gene mutations were detected in three unrelated individuals with early-onset colon cancer and less than 20 adenomatous colon polyps. A SMAD4 mutation was detected in a proband with gastric cancer at age 35, who was later found to harbor a polyp of varying pathology types, none of which were juvenile, by age 50. Another SMAD4 mutation was found in a proband diagnosed with colon cancer at age 30 with normal tumor MSI and IHC testing. A PTEN mutation was identified in a 65 year old individual with over 100 adenomatous colon polyps and no family history of cancer. An individual with 2-5 adenomatous polyps and colon cancer at age 39 was found to carry an APC mutation. A 16 year old with colon cancer and a reported heavy polyp load was found to carry two MSH6 mutations. A 36 year old unaffected individual with a MSH2 mutation was tested due to a paternal family history of sarcoma, endometrial, and late onset colon cancers. An intrinsic MSH2 mutation was found in a patient with endometrial cancer at age 48 but whose family history was not suggestive of Lynch Syndrome. A single exon PMS2 deletion was found in a proband with 2-5 colon polyps, colon cancer at age 49, and normal MSI and IHC tumor testing. A proband meeting Cowden syndrome clinical criteria was found to carry a PMS2 mutation. Since these results reveal that genotype-phenotype discrepancies clearly exist among individuals carrying mutations in well-known cancer syndrome genes, we encourage continued evaluation of the clinical features associated with each condition. We encourage the continued reporting of additional cases with uncharacteristic correlations in order to further expand upon the clinical criteria utilized in current diagnostic processes.


We have built a system facilitating sharing of unpublished data from whole exome (WES) and whole genome sequencing (WGS) studies. The system should help to speed up gene discovery by assisting researchers to find the critical ‘second case’. The main hurdle in current studies using WES/WGS is to find sufficient evidence to prove causality. Roughly studies give a yield of 1/3 proven causality, 1/3 likely causality and 1/3 unresolved cases. An obvious way to improve overall yield would be reaching out to colleagues worldwide to find additional positive or negative evidence for the candidate variants remaining. An important obstacle here is, for several reasons, the desire and demand to publish in peer-reviewed journals which works against pre-publication data sharing. The approach is built around the gene variant databases (LSDBs) by using version 3 of the LOVD platform (Leiden Open-source Variation Database, http://www.LOVD.nl) to collect and share information about genes, variants and phenotypes (diseases) and facilitate the analysis of exome and genome sequence data. Within the databases curated by us we have implemented a so-called VIP-status for both variants and phenotypes demanding specific attention. People can submit both phenotype descriptions and/or gene variants with the request to assign these a VIP status. VIP-phenotypes are those for which WES/WGS studies were performed but for which insufficient evidence was gathered to prove causality with variants in a specific gene. VIP-variants are those remaining after stringent filtering that could not be linked to nor excluded as causative in WES/WGS. The submitter decides whether the submission is named or anonymous. Others studying VIP-phenotypes or identifying VIP-variants are invited to contact the submitter to collaborate. For anonymous submissions the LOVD-curator will act as an intermediate to bring researchers into contact. We believe this option is effective regarding sharing data while at the same time protecting private interests and publication findings.

The Human Variome Project has granted LOVD the recommended system status for variant collection.
Background: With the completion of the Human Genome Project and the development of high throughput technologies, such as next-generation sequencing, the use of multiplex genetic testing is growing rapidly. The development of genetic cancer panels to assess multiple cancer risks represents one way in which multiplex testing is being applied clinically. There are a number of unique issues to consider when conducting genetic panel assessment for cancer risk assessment that differ from the traditional single-gene approach. Methods: To address an emerging need for multiplex sequencing panels, we set out to design a cancer panel comprised of genes that confer high or moderately increased risks for breast, ovarian, and colon cancers. Through this process we addressed the following issues: determination of genes to include in the panel, determination of risk estimates and how to convey them to the patient, review of surveillance and management guidelines for increased risk, the genetic counseling process, and return of results. Literature review, cancer gene databases, and existing cancer gene panels were used to inform the development of this panel. Results: Our cancer gene panel consists of 26 genes that confer high or moderate risks for breast, ovarian, and colon cancers, with some genes conferring risks for additional cancers. Many existing gene share molecular pathways such as the FANC-BRCA, CHEK2, and MMR pathways. Risks were generally found to cluster at 2 levels: Moderate risk genes, such as those in the FANC-BRCA and CHEK2 pathways for breast cancer, confer a 2-4 fold increased risk while high risk genes, such as CDH1 and PTEN, confer a 10-20 fold increased risk. Established guidelines for those with increased risk based on family history and clinical factors can be applied to those with comparable risk levels conferred by moderate penetrance genes. Conclusions: Cancer panel testing enhances the benefits of genetic risk assessment by 1) extending testing to a wider population beyond those who meet standard genetic testing criteria and 2) broadening the number of gene targets to assess risk, providing a more comprehensive risk assessment. By examining and integrating the data and tools currently available, we can maximize the clinical utility of cancer panel testing and identify the gaps in our knowledge that warrant further investigation.

2637T
Exome sequence of genetic disorders in consanguineal family. G. Altmann1, D. Ben Avraham2, B. Pode-Shakked2, Y. Anikster2. 1) Departments of Medicine and Genetics, Albert Einstein College of medicine, Bronx, NY, USA; 2) Metabolic Disease Unit, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer, Israel.

Consanguinity, is common in certain minority populations in Israel, naturally encompasses a greater risk for autosomal recessive disorders. In order to reveal the molecular mechanism and polygenic predisposition of two cases (same family) with autosomal recessive disorders (Osteogenesis imperfecta and Congenital Disorders of Glycosylation (CDG)), we employed an unbiased screening approach by using Exome sequence capture arrays followed by next generation sequencing (HiSeq2000). We Exome sequenced (due to the polygenetic anomaly of the investigated diseases) of 6 members of the family, the 2 cases kids their parents and one sibling of each of the parent, as the primary diagnosis of both disorders is of great significance for future family planning. Of the 154,432 sequence variants (available from the 6 samples) 28,487 variants passed the various QC filtering stages with 15 reads/variant. Of which 316 and 216 sequence variants, displayed heterozygosity in the unaffected samples and homozygosity in the affected ones (Osteogenesis imperfecta and CDG, respectively). While, forty missense mutation in the Osteogenesis imperfect case demonstrated the homozygosity pattern of inheritance, were the OR4 (olfactory receptor, family 4) gene's subtype exhibit the highest number of variants, among the CDG case 24 were of missense mutation with substantial prevalence of the FANCA gene (which encoded Fanconi anemia). In conclusion, Exome sequencing was demonstrated herein as a powerful tool for Molecular diagnosis of specific mutations and is therefore extremely necessary in such cases. Primary care physicians should be alert to the possibility of more than one autosomal recessive disorder in consanguineous families and evaluate them promptly with the use of such tool.

2639W
Recognition of Disease-Associated Alleles in the Reference Sequence is Critical for Accurate Disease-Risk Assessment through Genome Sequencing. G. Chandralilaka1, S. Garcia1, R. Chen1,2, M. Clark1, S. Chervitz1, D. Newburger1, H. Lam1, J. West1, R. Chen1. 1) Personalis, Inc., Menlo Park, CA; 2) Icahn School of Medicine, Mount Sinai, New York, NY.

The public reference genome sequence (GRCh37) contains minor alleles at >1 million positions. The presence of minor alleles in the reference negatively impacts both sequence read alignment and variant calling. For example, an individual who is homozygous for a minor allele present in the reference will not be reported as variant at that locus, resulting in failure to apply any medical interpretation relevant to that variant to the individual. On the other hand, individuals heterozygous or homozygous for major alleles absent in the reference are reported as variant at such loci, resulting in an increased burden of variant interpretation. To address the issue of rare alleles in the reference, we extended previous work (PMID 21935354) to create an enhanced human reference sequence. We revised 1.1 million positions in GRCh37 where the reference allele is the minor allele by frequency in four different populations. To assess the impact of this enhanced reference on disease-risk assessment, we interrogated GRCh37 for the presence of medically relevant minor alleles, identifying rare/minor alleles associated with Mendelian disease, pharmacogenetics, and complex disease. We found 38 variants in GRCh37 previously reported to be involved in Mendelian disease (HGMD designation DM/DM?) e.g. rs6784677 in BBS2 associated with Bardet-Biedl Syndrome and rs1529927 in SLC12A3 associated with Gitelman Syndrome and hypertension. An additional 217 variants designated disease-associated polymorphisms with functional evidence (DFP) were identified, e.g. the Factor V Leiden allele, rs9025. Furthermore, we identified 38 variants a奢华ious alleles (nonsense, frameshift, splice-site) in HGMD genes e.g. rs276936 in DSC3 and rs9959632 in PIGN. Whilst it is unlikely that all of these variants are pathogenic, they warrant in-depth review. Such variants would be completely missed in homozygous individuals and likely filtered out due to population frequency in heterozygous individuals due to recognition of the major allele as variant. In addition, we identified 77 variants with pharmacogenetic associations listed in PharmGKB, e.g. rs1954787 involved in citalopram response, and 985 variants identified with complex disease in our extended curated Disease Variant Database. The use of our enhanced human reference sequence facilitates more complete variant discovery–critical for accurate disease-, carrier- and pharmacogenetic-risk assessment through exome/ genome sequencing.

2638F

Objective: To assess the evidence behind the use of whole exome sequencing (WES) to identify genetic changes in cancer. Context: Next-generation sequencing technologies are advancing at a rapid rate, allowing the generation of a large amount of data in a relatively short period of time. Genetic changes in cancer are increasingly used for diagnosis and may guide treatment decisions. There are often more than 1, and sometimes many, different genes contributing to the clinical presentation of the disease. Given that the exome encompasses < 2% of the genome, variants affecting the cancer being studied may potentially be missed by WES; however, WES allows detection of 85% of disease-causing variants, thereby providing a less expensive method to detect variants when compared to whole genome sequencing. Nevertheless, the question remains: is there evidence that WES impacts patient outcomes in cancer? Methods: Proprietary methodology that combines the ACCE (Analytical validity; Clinical validity; Clinical utility; Ethical, legal and social implications) model for genetic test evaluations with internationally accepted health technology assessment methodology was used. Conclusions were based on peer-reviewed published studies of > 10 patients. Results: Whole WES has been conducted in > 10 patients for a number of different cancers (e.g. colon, prostate, and ovarian cancer), breast cancer has been evaluated most extensively (7 studies to date). Studies evaluating somatic alterations showed high intratumor and intertumor heterogeneity. In addition, both novel and previously implicated variants were identified, with varying frequencies based on the breast cancer subtype (e.g., luminal A or B, basal-like, or HER2-enriched). To date, only 2 studies with > 10 individuals (with breast cancer or ovarian cancer) have shown potential for clinical utility of WES, whereby variants identified through WES may determine response to drug treatment. Conclusions: Despite evidence for clinical validity of WES in cancers, clinical utility is very limited and needs to be further evaluated in large clinical studies. In addition, WES results may have ethical implications (e.g., incidental findings); while recent recommendations suggest that such findings must be provided to physicians/ patients, the appropriate way to communicate this information is not completely clear.
2640W
Autism associated with an Xq12 deletion involving the gene OPHN1 - importance of pursuing a genetic etiology for an Autism Spectrum Disorder (ASD).

E. Etias. Dept Genetics & Pediatrics, Children's Hosp Colorado, Aurora, CO.

Introduction: ASD's are prevalent, occurring at a rate of 1:88 children per recent CDC report. Many children with an ASD never undergo genetic evaluation. However, current technology allows the determination of a more definitive underlying genetic diagnosis for children with an ASD, especially if accompanied by intellectual disability (ID).

Case Report: A 9 year old boy presented for genetic evaluation of an ASD, and severe ID. Past medical history was notable for strabismus requiring surgical repair and a seizure disorder requiring medication. His exam revealed mildly dysmorphic features, including down-slanting palpebral fissures and a prominent chin. Head MRI was notable for inferior cerebellar hypoplasia, a Dandy-Walker variant and mild ventriculomegaly. Previous genetic testing included normal karyotype and negative Fragile X testing. Family history was notable for a normal brother, and healthy parents.

Genetic Testing: A 191 kb region of loss involving chromosome Xq12 was detected on a Cytochip 180K Oligo microarray panel, which partially overlapped the OPHN1 gene. Similar patients have been reported with either deletions of Xq12, or molecular mutations of the OPHN1 gene. Oligophrenin-1 (OPHN1) encodes a Rho-GTPase-activating protein, the first Rho-linked protein identified in patients with X-linked ID. Rho-GTPase proteins are critical for normal neuronal development and function, particularly regarding dendritic growth and genes affecting glutamate receptors.

Conclusions: This patient had a constellation of problems including strabismus, seizures, and cerebellar anomalies, markedly similar to other patients who have been reported with this rare chromosome deletion syndrome affecting the OPHN1 gene, and is further confirmation of the importance of this gene in CNS morphogenesis. Updated genetic evaluations can be revealing for patients with ASD, especially for those presenting with ID, dysmorphic features, and anomalies. Establishing a definite genetic diagnosis also allows appropriate genetic counseling for this X-linked ID disorder.

2642F
A Balanced Chromosome Translocation Reveals Involvement of a Predicted Lipase in Weight Gain, Hearing Loss and Tumor Suppression.

B.B. Currall1,2, K.E. Wong1, N.G. Robertson1, T. Hoyos3,2, A.E. Hickox4, Y. Yin5,2, B.J. Quade1,2, M.C. Liberman7,2, E.C. Liao3,2, C.C. Morton1,2,1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Massachusetts General Hospital, Boston, MA; 4) Purdue University, West Lafayette, IN; 5) Massachusetts Eye and Ear Infirmary, Boston, MA.

The Developmental Genome Anatomy Project (DGAP; www.dgap.harvard.edu) systematically examines subjects with balanced chromosomal rearrangements to identify genes involved in congenital disorders. DGAP056 is an archetypal case with a constellation of symptoms including profound congenital hearing loss, early onset prostate cancer, mild craniofacial abnormalities (coloboma, exotropia, blepharophimosis, and low-set posteriorly rotated ears), mitral valve prolapse and hypospadias. Genomic analyses of DGAP056, using karyotyping, fluorescence in situ hybridization (FISH) and genomic sequencing, revealed a balanced translocation involving chromosomes 2 and 13, t(2;13)(p24.1;q22.3)dn, which disrupts a poorly annotated gene designated C2orf43. Linkage and genome-wide studies have associated SNPs in and around C2orf43 with prostate cancer, coronary heart disease and other phenotypes associated with lipid deregulation. Phylogeny and tertiary structure modeling indicate that C2orf43 encodes a conserved protein related to the alpha/beta hydrolase clan of proteins and likely functions as a serine-based ester hydrolase involved in lipid metabolism. In animal models, C2orf43 orthologs are expressed in a variety of tissues including adipose tissues, inner ear and prostate. Mouse knockout models have increased weight, high frequency hearing loss and increased rates of tumors (particularly in the prostate). C2orf43 expression is also down-regulated in human prostate tumors. While it is yet unclear the biochemical role of C2orf43, this gene appears to be necessary for proper lipid metabolism and, when disrupted, leads to a variety of diseases including obesity, hearing loss and prostate cancer. DGAP056 demonstrates the power of the DGAP approach using de novo genetic lesions, well characterized phenotypes and convergent data to annotate clinically important regions of the genome.

2641T
CELL MALIGNITATION ASSOCIATED TO CHROMOSOME TRANSLocations. CLINICAL MANIFESTATIONS IN TWO PEDIATRIC PATIENTS

46,XY,t(1;4)(q11q11) AND 46,XY,t(6;9)(p21;q34), J. Aparicio1,3, M.L. Hurtado4, S. Chatelain1,4. 1) Dept Gen; 2) Cytogenetics, Hosp para el Nino Pobiano, Puebla; 3) Estomatology, benemerita Universidad Autonoma de Puebla; 4) Biotechnology, Universidad Autonoma Metropolitana Mexico DF.

Chromosomal aberrations are considered alterations in the chromosome number or structure. They are mainly considered due to gametogenesis inborn error or during the zygote first cellular divisions. Among 4617 chromosome studies performed during 19 years (from 1992 to 2011), at Hospital Para El Niño Pobiano in México, 34.6% (1596 patients) had chromosomal alterations. Among these studies population, 0.23% (11) chromosome translocations were observed. From this data, two male pediatric patients are described, with 1;4 and 6;9 chromosome translocations. Chromosome changes are classified as structural or numeric alterations respectively, and abnormal cell development has been associated with these two specific chromosomal translocations. Both cases were described in this study analyzing their hematological, clinical features, medical treatments and prognosis.
2643W
Genitourinary Defects Associated with Genomic Deletions in 2p15 Involving OTX1. C. Jorge1,2, J. Rosenfeld3, N. Wilken1, V. Vangapandu1, A. Sahin4,5, D. Pham2, C. Carvalho4, A. Bandholz6, A. Miller7, D. Weaver7, B. Burton8, D. Babu9, J. Bamforth9, T. Wilks10, D. Flynn11, E. Roeder12, S. Cheung13, J. Lupski14, D. Lamb12,13. 1) Center for Reproductive Medicine, Baylor College of Medicine, Houston, TX; 2) Scott Department of Urology, Baylor College of Medicine, Houston, TX; 3) Department of Molecular and Cell Biology, Baylor College of Medicine, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 7) Department of Medical and Molecular Genetics Indiana University School of Medicine, Indianapolis, IN; 8) Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 9) University of Alberta, Edmonton, Alberta, Canada; 10) Madigan Army Medical Center Department of Pediatrics, Tacoma, WA; 11) Department of Children's Endocrinology, St. Luke's Children's Specialty Center, Boise, ID; 12) Department of Pediatrics University of Texas Health Science Center at San Antonio, San Antonio, TX.

Normal development of the genitourinary (GU) tract is a highly complex process that frequently goes awry, causing malformations of GU structures. In male children the most frequent congenital anomalies are GU defects such as cryptorchidism (1% to 4% of full term newborns), hypospadias (1%), micropenis (0.35%) and vesicoureteral reflux (VUR) (1%). Other congenital urologic malformations, such as bladder-exstrophy-epispadias (BEE) (1:47000), occur less frequently but significantly impact patients' lives. Array comparative genomic hybridization (aCGH) identified seven individuals with deletions in the 2p15 region (66.0kb-5.6Mb). These deletions encompass transcription factor orthodenticle-homolog-1 (OTX1) gene. Subject 1 (with BEE and VUR) had the smallest deletion (66kb), encompassing only OTX1, and was identified among 30 BEE patients screened by aCGH. Male subjects 2-5 were identified among 30,183 subjects submitted to Signature Genomics for clinical aCGH testing, and male subjects 6 and 7 were identified among 18,734 subjects analyzed at Baylor Genetic Laboratories for clinical aCGH testing. Subjects 2-7 had large de novo CNVs (2.39-6.31 Mb).

Involving OTX1.

2644T
6q24.3-q25.1 deletion syndrome. Y. Nishi1, M. Tominga2, H. Ueda2, Y. Kuroda1, I. Ohashi1, T. Saito2, J. Nagai3, K. Kurosawa4. 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Department of Cardiology, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Clinical Laboratory, Kanagawa Children's Medical Center, Yokohama, Japan; 4) Japan Science and Technology Agency, CREST.

6q24-25 deletion syndrome is a rare and newly recognized syndrome characterized by congenital heart defects, growth retardation, and variable degree of intellectual disability. The phenotype varies with the range of deleted chromosomal segments and the genes involved. Nakayama et al. (2009) reported three patients with paternal deletion of 6q24.4. Two of the patients shared a 2.5 Mb region of overlap and similar facial features including a triangular face, frontal bossing, short and up-slaning palpebral fissures, asymmetry of upper eyelids, shallow orbits, and long and flat philtrum. The cardiac defects include the anomaly in the outflow tract and atroventricular defect. Thienpont et al. (2010) demonstrated that the cardiac defect of the syndrome was caused by haploinsufficiency of TAB2 gene harbored on 6q25.1. Here we present an additional case of 13-year-old girl with the 6q24.3-q25.1 deletion. She was born to nonconsanguineous healthy parents at 40 weeks of gestational age after uneventful pregnancy. Her birth weight was 2780 g. Recurrent episodes of pneumonia were noted during the infantile period. Her developmental milestones were delayed; head control at 4 months, walking alone at 1 year and 6 months, speaking comprehensive words at 2 years and 6 months. At age of 6 years, G-banded chromosome analysis was performed due to her mild developmental delay, but the result was normal karyotype. Her intelligence quotient was 51-75 at age of 12 years. She had short stature due to a deficiency of growth hormone and delayed puberty started at her age of 7 years. Since the age of 8 years, she had precocious puberty and treated with GnRH antagonist. At age of 12 years, atrial premature contraction and severe mitral regurgitation was noted. Because of uncontrollable attacks of arrhythmia, she had catheter ablation therapy repeatedly. To elucidate the cause of these clinical episodes, cytogenetic microarray was performed. The results showed interstitial deletion of 4.2 Mb from 6q24.3-q25.1. The present patient shared several features including the facial anomaly, the degree of intellectual disability, growth phenotype, and characteristic cardiac defect. Together with these results and previous reports, the deletion of 6q24.3-q25.1 represents distinctive microdeletion syndrome caused by haploinsufficiency of the genes in the interval. Further analysis is required to elucidate the phenotype-genotype correlation in the syndrome.

2645F
An infant with 49.XXXX syndrome and congenital cataract. C. Vinkler1, A. Ben Sasson2, A. Singer3. 1) Inst Med Genetics, Wolfson Med Ctr, Holon, Israel; 2) Child Developmental Center, Maccabi Health Service; 3) Inst Med.- Genetics Barzilai Medical Center , Ashkelon, Israel.

49.XXXX syndrome is one of the rarest sex chromosome abnormalities in humans, showing an incidence of 1:100,000. It is described to be the most severe variant of Klinefelter syndrome however it is categorized as a separate syndrome by many authors. Clinical features of the syndrome are: coarse face, microcephaly, distinct dysmorphic features, short stature and an IQ ranging between 20 to 60 points. Recently, increased rates of brain anomalies were also described in this group of patients. We present a patient with 49.XXXX syndrome and congenital cataract. The proband is a 2y and 3m old boy referred to our clinic because of severe hypotonia, developmental delay and dysmorphic features. This male patient was born at term to healthy parents. There is no description of cataracts in the family. Pregnancy and delivery were normal with a birth weight of 2.7 kg. Right after birth hypotonia and dysmorphic features were noticed. Cardiac echo revealed small PFO and PDA Eye examination revealed Lt anterior polar cataract of 1 mm and another posterior subcapsular cataract. Brain ultrasound was normal. Chromosome karyotyping showed a karyotype of 49.XXXX. On examination at the age of 2y and 3m his head circumference was 48cm (3rd centile) height was 75cm (-4 SD) and weight 9kg (-3.5 SD). He has a coarse facial appearance with hypertelorism, epicanthal folds, upslanting palpebral fissures and a broad nasal bridge. He has small penis and small retractile testes. He is diagnosed with global development delay. This is the first report of congenital cataract in a patient with 49.XXXX syndrome. Congenital cataract has been described previously, in one case of 46.XXY syndrome. The molecular basis for the combination of congenital cataract in sex chromosome polysomy is not known. It had been previously suggested that it may be explained by an increased BCCR gene dosage. Mapped to patient's extra X chromosome X, mutations of the BCCR gene can result in cataracts, and other ocular abnormalities. X inactivation may explain the variable phenotype and rare expression of cataracts in these patients. Further investigation in model systems is needed to verify this hypothesis. Eye examination is recommended in all patients with sex chromosome polysomy.
2646W
Aiding the interpretation of CNV and sequence variation in DECIPHER using the Genovverse genome browser. E. Bragin, E.A. Chatzimichali, G.J. Swaminathan, A.P. Bevan, C.F. Wright, M.E. Hurles, H.V. Firth, Wellcome Trust Sanger Institute, Wellcome Trust Campus, Hinxton, United Kingdom.

DECIPHER (https://decipher.sanger.ac.uk) is a web-accessible database and consortium that facilitates the identification and interpretation of genomic variation in patients with developmental disorders. Over 250 academic departments in genetic medicine contribute phenotype-linked variation data into the DECIPHER database for analysis and interpretation. Following informed consent, shared anonymized patient data enables the identification of clusters of patients with similar phenotype-linked genomic findings and encourages collaboration and contact between member centers. DECIPHER also facilitates contact between external users and consortium members making it an invaluable collaborative resource for genomic research and clinical diagnosis. Driven by decreasing costs and improved technologies, sequencing is now increasingly being used in clinical diagnosis alongside arrayCGH. In order to facilitate the combined analysis and interpretation of phenotype-linked copy-number (CNV) and/or sequence variation we have extended and improved DECIPHER to encompass all forms of genomic variation. New features include informative summary and gene tables, variant-effect predictions as well as a purpose-built genome browser (Genovverse: http://genovverse.org) to visualize and interpret copy-number and sequence variation. Genovverse was developed in the Wellcome Trust Sanger Institute in collaboration with Ensembl. The Genovverse browser utilizes modern web technologies to visualize data ‘on-the-fly’. Some of the salient features include customizable views, interactive scroll and zoom, as well as visualization of all filtering steps by ‘drill down’. Genovverse is currently the visualization engine behind DECIPHER and Ensembl. The design of Genovverse allows easy integration into any website and connection with various data formats and sources. In our presentation we demonstrate our implementation of all recent developments in DECIPHER and Ensembl. The integration of sequence and CNV data, managed data access, inclusion of research data and the new genome browser.

2647T
The Diagnostic Yield of Chromosomal Microarray Analysis in a Large Multidisciplinary Craniofacial Clinic. K. Dipple1, J. Peredo2, J.P. Bradley2, R. Jarrahy2, F. Quintero-Rivera2. 1) Dept Hum Gen, Gonda 55068, Univ California, Los Angeles, Los Angeles, CA, 2) Dept Surgery, Division of Plastic and Reconstructive Surgery, Univ California, Los Angeles, Los Angeles, CA; 3) Dept Pathology and Laboratory Medicine, Univ California, Los Angeles, Los Angeles, CA.

We report on 100 children tested in the UCLA Craniofacial Clinic using chromosomal microarray in an effort to locate significant areas of the genome and candidate genes associated with craniofacial anomalies. Abnormalities included isolated defects (cleft lip/palate, cleft palate alone, hemifacial microsomia, microtia, Pierre Robin sequence and Tessier clefts), known chromosomal microdeletions or microduplications, common disorders like FRAXE (without CHD7 mutations or deletions), Nager, OAVID/Goldenhar) and individuals with the above isolated defects plus multiple abnormalities, developmental delay and/or intellectual disability. We excluded those patients who had previously identified chromosomal abnormalities by standard G-banded karyotyping. Patients were tested from September 2007 to May 2013 using BAC arrays and later high density oligonucleotides (with and without single nucleotide polymorphism, SNP) arrays once that platform became available. The clinical positive rate of microarray testing in the UCLA Craniofacial Clinic was 10% (n=10/100). Four patients had large regions of homozygosity (ROH) that ranged from 10 Mb to 389 Mb, (one case was 2nd degree relationship). One patient had a partial tetrasomy/deletion where the gain was not associated with a supernumerary chromosome (at the Xp11.21, being consistent with Turner syndrome. (Jarrahy, J. Dipple, 2013). Another patient had a 5q35.2-q35.3 microdeletion associated with Sotos syndrome. (Peredo J, et al., J. Craniofac J. 2012). We did not find any overlap with previously reported candidate genes for isolated cleft lip/palate or Goldenhar syndrome. Our experience in this population shows that those children with cleft palate and hemifacial microsomia plus multiple abnormalities (congenital heart defects, encephalopathies, seizures, muscle tone abnormalities) and developmental delay are the most likely to have deletions/duplications that can be detected using CGH array. Whole genome sequencing (WGS) in patients whose craniofacial abnormalities were isolated.

2648F
Effects of up-regulation of the SHH pathway on Ts65Dn, a mouse model of Down Syndrome. T. Dutka1, N. Singh1, J.T. Richtsmeier1, R. Reeves1,2. 1) Institute Of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Department of Anthropology, Pennsylvania State University, University Park, PA; 3) Department of Physiology, Johns Hopkins School of Medicine, Baltimore, MD.

Down Syndrome (DS) is caused by a triplication of human chromosome 21 (Hsa21). DS consists of a constellation of over 80 phenotypes. Three characteristics found in all individuals with DS are some level of craniofacial dysmorphology, brain structural and functional changes, and cognitive disability. Ts65Dn, a model system for DS, was generated in mice by introducing an extra chromosome consisting mostly of the distal portion of mouse chromosome 16 (Mmu16), a region orthologous to Hsa21. Previous work has demonstrated that Ts65Dn recapitulates aspects of the craniofacial defects, brain dysmorphology, and intellectual disability seen in DS. Moreover, some aspects of each of these 3 traits have been linked to a reduced response to Sonic Hedgehog (SHH). If all trisomic cells show a similarly reduced response to SHH, then up-regulation of the pathway in affected cells might ameliorate the phenotypes in multiple tissues. To investigate this hypothesis, Ts65Dn mice were crossed with Ptch1tm1Mps/+. Mice in which the canonical SHH pathway is up-regulated in every SHH-responsive cell through the loss of function of one Ptch1 allele; this should affect every cell that signals through the cranial HH pathway throughout development. Ts65Dn; Ptch1tm1Mps/++; mice were compared to Eu; Ptch1tm1Mps/++; Eu; Ptch1tm1++; and Ts65Dn; Ptch1tm1++ for craniofacial, behavioral and brain phenotypes. Morphometric assessment of the skull demonstrated specific effects on craniofacial morphology. Evaluation of motor learning was accomplished using an increasing speed rotarod test (walking speed drop). Once one is current with the visualization engine behind DECIPHER and Ensembl. The design of Genovverse allows easy integration into any website and connection with various data formats and sources. In our presentation we demonstrate our implementation of all recent developments in DECIPHER. Our experience in this population shows that those children with cleft palate and hemifacial microsomia plus multiple abnormalities (congenital heart defects, encephalopathies, seizures, muscle tone abnormalities) and developmental delay are the most likely to have deletions/duplications that can be detected using CGH array. Whole genome sequencing (WGS) in patients whose craniofacial abnormalities were isolated.

2649W
46, X, (p11.2) Turner syndrome patient with severe keratoconus. L. Gabriel1, 2. L. Junior1, T. Oliveira1, I. Silva1, L. Chaves1, L. Sousa1, L. Elias1, L. Mendonça1, R. Filho1, L. Lavigne1,2, J. Filho1, J. Jaime1, M. Avila2, 3. 1) Ophthalmology, CEROFC-UFG, Goiania, GO, Brazil; 2) Bra- zilian Center for Eye Surgery, BCES, Goiania, GO, Brazil; 3) APAE, Anapolis, GO, Brazil; 4) Centro de Diagnostico por Imagem, CDI, Goiania, GO, Brazil; 5) Hospital Geral Roberto Santos, Salvador, BA, Brazil. A 19-year-old female with amenorrhea, short stature, and severe keratoconus on both eyes was investigated for the diagnostic hypothesis of Turner syndrome. We examined and compared C.H.F. (Ear, nose, mouth) and cataract in Turner syndrome. (DHEA), and a G-banding karyotype showing the X chromosome partial deletion 46, X, (p11.2). Confirming our suspicion of Turner syndrome.

2650T
Heterotaxy in a woman with mosaic Turner syndrome. P. Kannan1, A.E. Lin1, N.S. Scott1, S. Sahai1. 1) Medical Genetics, MassGeneral Hospital for Children, Boston, MA; 2) Department of Cardiology, Massachusetts General Hospital, Boston, MA.

BACKGROUND: Heterotaxy refers to an abnormal arrangement of the thoracic and abdominal organs due to a failure to establish left-right asymmetry during embryogenesis. Important clinical findings include complex cardiovascular malformations, abnormal inferior vena cava (IVC) relationship to the spine, spleen defects, transverse liver, and intestinal malrotation. Heterotaxy is genetically heterogeneous. Not all causative genes are known. Chromosomal abnormalities are uncommon. Turner syndrome is a chromosomal disorder occurring in 1 in 2500 live female births. Short stature, ovarian dysgenesis and heart defects are common findings. Turner syndrome is caused when an entire X chromosome is missing (45,X) or is structurally abnormal. In mosaic Turner syndrome 45,X/46,XX, a milder clinical presentation in typically observed. CASE: We present a 29 year old woman with a history of Total Anomalous Pulmonary Venous return into coronary sinus status post surgical repair in infancy, left sided IVC, polypslena, left of midline liver and intestinal malrotation. Her karyotype is 45,X/46,XX(27). Sequencing of the 10 known genes associated with heterotaxy did not reveal any variants. Heterotaxy in Turner syndrome is extremely rare and has been reported in two cases previously; our case is the only one in which known heterotaxy genes were analyzed.
2651F
Dilated cardiomyopathy in a patient with Pallister Killian Syndrome while on a ketogenic diet. J. Lazier1, J. Harder2, M.A. Thomas1, 1) Department of Medical Genetics, Alberta Children's Hospital, Calgary, Alberta, Canada; 2) Department of Pediatrics, Alberta Children's Hospital, Calgary, Alberta, Canada; 3) Department of Pediatric Cardiology, Alberta Children's Hospital, Calgary, Alberta, Canada.

Pallister Killian Syndrome (PKS) is caused by mosaic tetrasomy of chromosome 12p. It is characterized by multiple congenital anomalies, characteristic craniofacial anomalies, and intellectual disability. Seizures are common and often poorly controlled with medication. Up to 40% of patients with PKS have congenital heart anomalies, including atrial septal defects, patent ductus arteriosus and bicuspid aortic valves. Cardiomyopathy is a rare association, with only one reported case of hypertrophic cardiomyopathy. There are no previously reported PKS patients with dilated cardiomyopathy as seen in our patient. Vogel et al. (2009) reported a case of a combined hypertrophic/ dilated cardiomyopathy reported in a patient with hexosam 12p, and Parasuraman et al. (2011) reported a case of severe dilated cardiomyopathy reported prenatally in a fetus with mosaic trisomy 12. These cases suggest that upregulation of genes in this area may contribute to dilated cardiomyopathy.

Ketogenic diets are commonly used to treat children with severe refractory epilepsy, such as seen in PKS. While they are generally considered safe, long term use has been associated with metabolic changes, including selenium deficiency, which is present in up to 20% of patients on this diet. Rare cases of dilated cardiomyopathy and long QT syndrome have been seen in these selenium deficient patients.

We present the case of a four year old girl with PKS who presented with a long QT syndrome and acute dilated cardiomyopathy. An echocardiogram done 15 months previously had been normal. At the time of presentation, she was on a ketogenic diet for medically refractory epilepsy and had low selenium levels. Despite cessation of the ketogenic diet, medical treatment of her heart failure, and initial improvement of her cardiomyopathy, the patient died from congestive heart failure two months after the initial presentation. Dilated cardiomyopathy has not been previously reported in PKS. Given that there is a low risk of cardiomyopathy in all patients who are selenium deficient, we have not pursued a genetic evaluation in this case. Given the low risk of cardiomyopathy in all patients who are selenium deficient, we have not pursued a genetic evaluation in this case. However, the patient's family has been offered genetic counseling.

2652W
Clinical and cytogenomic evaluation in two siblings with an 8.5 Mb 6q24.2q25.2 deletion inherited from a paternal insertion. M. I. Melaragna1, S.S. Takeno1, M. Megliavacca1, A.L. Pilla1, N.L.M. Sobreira1, C.B. Melo1, V.A. Meloni1, 1) Genetics Division, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; 2) Mackusick Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, USA; 3) Psychology Department, Universidade Federal de Sao Paulo, Sao Paulo, Brazil.

Clinical and cytogenomic evaluation in two siblings with an 8.5 Mb 6q24.2q25.2 deletion inherited from a paternal insertion. M. I. Melaragna1, S.S. Takeno1, M. Megliavacca1, A.L. Pilla1, N.L.M. Sobreira1, C.B. Melo1, V.A. Meloni1, 1) Genetics Division, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; 2) Mackusick Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, USA; 3) Psychology Department, Universidade Federal de Sao Paulo, Sao Paulo, Brazil.

Given that there is a low risk of cardiomyopathy in all patients who are selenium deficient, we have not pursued a genetic evaluation in this case. However, the patient's family has been offered genetic counseling.

2653T

We report the case of a 9-year-old boy with a duplication within band 12q13.12 which is de novo and not previously described. He presented for genetics evaluation for developmental delay and dysmorphic features. His medical history was otherwise significant for small size, microcephaly, and undescended testis. An unknown congenital cardiac condition reportedly resolved in infancy, and he is now followed by cardiology for mildly increased tricuspid and total cholesterol. He had mild motor delays, walking around 18 months, but more significant delays in speech. He said his first words after his second birthday and did not use sentences until age four. At the time of his visit he was in the third grade but doing what his mother described as kindergarten-level work. On review of family history his mother reported removal of a unilateral cataract at age 5 thought to be secondary to toxoplasmosis. The remainder of the family history was significant only for a distant cousin on each parent's side of the family thought to have diagnoses on the autism spectrum. On physical exam his height and weight were at the 3rd percentile and his head circumference was less than the 3rd percentile and 50th percentile for 3 years of age. His face was significant for upplanting palpebral fissures without epicantthic folds, double cowlicks and an overall impression of mild facial asymmetry and mild micrognathia. Previous genetic workup at that time included normal karyotype, FISH for SNP and testing for Fragile X syndrome. A whole-genome oligonucleotide array CGH with SNP was done which revealed a de novo duplication of at least 417 kb within band 12q13.12. This duplication has not been previously described to vary in its entirety in the general population. It contains 4 genes with known clinical associations, including TUBA1A and ML2 associated with neurodevelopmental phenotypes. Heterozygous mutations in these two genes are associated with lissencephaly type 3 and Fukuyama muscle disease type 1, respectively. The clinical consequences of having three copies of these genes are not known. Based on the de novo occurrence the laboratory has now classified this as a positive result representing a possible new microduplication syndrome.

2654F
DIAGNOSTIC AND MANAGEMENT CHALLENGES OF GENETIC DISEASES IN RWANDA. L. Mutesa1,2, A. Uwineza1,2, J. Hitayezu1, S. Muronkwere1, E. K. Rusingiza2, T. Tuyisenge1, R. Tekeli2, J. Mucumbitsi4, N. Muganga1, A. C. Helli2, M. Jamari1, V. Bours3. 1) Medical Genetics, National University of Rwanda, Huye, Huye-Butare Southern Province, Rwanda; 2) Center for Human Genetics, University of Liege, Belgium; 3) Department of Pediatrics, Kigali University Teaching Hospital, Rwanda; 4) Department of Pediatrics, King Faysal Hospital, Kigali, Rwanda.

Background: Genetic diseases refer to genetic disorders caused by defects or abnormalities in chromosomes or genes. Chromosome disorders include numerical anomalies (e.g. monosomies, trisomies, etc) and chromosomal rearrangements mainly characterized by unbalanced translocations, inversions, deletions, duplications and translocations. In addition, microarray defects affecting gene expression represent a high proportion of monogenic or polygenic genetic diseases. Methods: In the present study, we conducted a survey aiming at assessing clinical aspects and genetic diagnostic patterns of Rwandan patients presenting with clinical phenotypes suggestive of genetic disease. This survey was a compilation of several genetic studies so far conducted within a six-year-period starting from 2006 to 2012. Most of patients were selected based on clinical features suggesting of a specific genetic disease. Cytogenetic and genetic studies include: karyotyping, Fluorescent In situ Hybridization (FISH) analysis were performed in the majority of these patients. In addition, molecular tests such as Polymerase Chain Reaction (PCR), gene sequencing or Multiplex Ligand Probe Amplification (MLPA) analysis and CGH array were done.

Discussion: In some situations, genetic counseling and testing should be mandatory, especially when effective treatment is available. These findings indicate that there is a need for the establishment of a well-equipped Genetic Laboratory in Rwanda that could screen for monogenic or chromosomal disorders. Furthermore, given the fact that genetic counseling and testing should be mandatory, especially when effective treatment is available. These findings indicate that there is a need for the establishment of a well-equipped Genetic Laboratory in Rwanda that could screen for monogenic or chromosomal disorders.
2655W  
**Neonatal Management of Trisomy 13: Clinical Details of 12 Patients Receiving Intensive Treatment.** E. Nishi1,2, M. Takasugi1, T. Hiroma1, T. Nakamura1, Y. Fukushima1, T. Kosho1,2,1) Department of Medical Genetics, Nagano Children’s Hospital, Azumino, Japan; 2) Department of Medical Genetics, Shinsu University School of Medicine, Matsumoto, Japan; 3) Department of Neonatology, Nagano Children’s Hospital, Azumino, Nagano, Japan.

Trisomy 13 is one of the most common autosomal trisomy syndrome, characterized by multiple congenital anomalies, severe developmental delay, and a short life span with a 1-year survival rate as 5-10 % and the median survival time as 7 days through population-based studies. Management of neonates with trisomy 13 is controversial, palliative care or intensive treatment, supposedly due to the lack of precise clinical information of the syndrome especially on efficacy of treatment. To delineate the natural history of trisomy 13 managed under intensive treatment, we reviewed detailed clinical data of 12 patients with full trisomy 13 admitted to the neonatal intensive care unit of Nagano Children’s Hospital, providing intensive treatment to those with trisomy 13, from 2002 to 2012. Two patients had prenatal karyotyping through amniocentesis. Five were delivered by a cesarean. Major clinical findings included congenital heart defects (92%), cryptorchidism (83% of male), cleft lip or cleft palate (75%), polycystic kidney (20%), tracheomalacia or laryngomalacia (58%), prolonged hypoglycemia (50%), gastroesophageal reflux (50%), and biliary tract disease (50%). 10 patients (83%) received resuscitation by intubation. Mechanical ventilation was required by 10 (83%), two (20%) of whom were extubated and six of whom needed tracheostomy. Surgical operations included tracheostomy, abdominal operation for umbilical hernia, ileostomy, hepatic portenterostomy, plastic operation for cleft lip, and operation for craniosynostosis. Perinatal death occurred on 10 (83%), of the patients. Enteral feeding was accomplished in all, four of whom were fed orally. Three patients could be discharged home and are still alive. Common final modes of death were congenital heart defects and heart failure (67%). The survival rate at age 1 day, 1 week, 1 month, and 1 year was 100%, 50%, 100%, and 92%, and 17%, respectively. Median survival time was 229.5 days (range, 22-2334). These data have a resemblance to the data about the patients with trisomy 18 in our hospital. (Kosho et al. 2006)

2656T  
**Options to families of patients with trisomy 13.** T. Nakamura1, Y. Fukushima1, T. Kosho1,2,1) Department of Medical Genetics, Nagano Children’s Hospital, Azumino, Japan; 2) Department of Neonatology, Nagano Children’s Hospital, Azumino, Japan.

The current study, though the sample size is small, has demonstrated improved survival through neonatal intensive treatment, which helps to inform families of patients with this rare syndrome. We report the clinical data of 12 patients with full trisomy 13 admitted to the neonatal intensive care unit of Nagano Children’s Hospital, providing intensive treatment to those with trisomy 13, from 2002 to 2012. Two patients had prenatal karyotyping through amniocentesis. Five were delivered by a cesarean. Major clinical findings included congenital heart defects (92%), cryptorchidism (83% of male), cleft lip or cleft palate (75%), polydactyly (67%), tracheoesophageal fistula (50%), and biliary tract disease (50%). 10 patients (83%) received resuscitation by intubation. Mechanical ventilation was required by 10 (83%), two (20%) of whom were extubated and six of whom needed tracheostomy. Surgical operations included tracheostomy, abdominal operation for umbilical hernia, ileostomy, hepatic portenterostomy, plastic operation for cleft lip, and operation for craniosynostosis. Perinatal death occurred on 10 (83%), of the patients. Enteral feeding was accomplished in all, four of whom were fed orally. Three patients could be discharged home and are still alive. Common final modes of death were congenital heart defects and heart failure (67%). The survival rate at age 1 day, 1 week, 1 month, and 1 year was 100%, 50%, 100%, and 92%, and 17%, respectively. Median survival time was 229.5 days (range, 22-2334). These data have a resemblance to the data about the patients with trisomy 18 in our hospital. (Kosho et al. 2006)

2657F  
**National Institutes of Health Activities in Down Syndrome Research: From Creation of a Consortium to Support of a Patient Registry, DS-Connect, to Patient Crossroads.** Y. Kuroda1, Y. Fukushima1, T. Kosho1,2,1) Department of Medical Genetics, Nagano Children’s Hospital, Azumino, Japan; 2) Department of Medical Genetics, Shinsu University School of Medicine, Matsumoto, Japan; 3) Department of Neonatology, Nagano Children’s Hospital, Azumino, Nagano, Japan.

Down syndrome (DS) is one of the most common genetic causes of intellectual disability worldwide. The National Institutes of Health (NIH) has a longstanding program of research in DS to improve understanding of this chromosomal disorder, with the goal of developing effective treatments to improve quality of life for those with DS. In 2007, the Trans-NIH Down Syndrome Clinical Research Network (DS-CONNECT), led by the NICHD, was launched to identify, characterize, and define the best care practices for DS. In September 2012, NICHHD awarded a contract to create DS-Connect: The NIH Down Syndrome Consortium Registry to facilitate information sharing among families, persons with DS, researchers, and parent groups. Family members and those with DS will be able to enter contact and health information into an online, confidential, and secure database. A researcher and professional portal will allow evaluation of aggregated de-identified data. If a participant is eligible for a research study, Registry staff will invite them to contact the investigator for potential participation in the study. DS-Connect is planned for launch within the coming year. These activities highlight progress and remaining challenges in DS research. With new therapeutics being developed to treat the cognitive deficits in DS, the need for these resources is timely.

2658W  
**Phenotype correlation of a patient with a large 16q23.2 to 16q24.3 duplication and a patient with a 16q23.3 to 16q24.3 duplication and small 16p13.3 deletion.** J. Riche1,2, J. McGowan-Jordan1,2.

1) Genetics Department, CHEO, Ottawa, Canada; 2) University of Ottawa, Department of Pediatrics, Ottawa, Canada.

Patients with large duplications involving the long arm of chromosome 16 are extremely rare. A limited number of patients with large duplications of chromosome 16 with additional rearrangements are described in the literature; descriptions in the literature of patients with isolated large interstitial duplications are rare. We present a girl who has a de novo pure duplication of the 16q23.3 to 16q24.3 (16q23.3q24.3 dup), who was not included in the Clinical and Genetic Analysis of Large Chromosome Rearrangements (CAGAR) database. Case 2: The patient is a 16 year old female. She showed a large terminal 16q duplication of 16q23.2 to 16q24.3, as well as a 16p13.3 terminal microdeletion. In conclusion, the dysmorphic features and cognitive impairment. There were also concerns about recurrent upper respiratory infections and about the child’s ability to tolerate vaccination. Genomic SNP microarray showed a 9.8 Mb duplication of 16q23.2 to the terminus at 16q24.3. Case 2: The patient is a 16 year old female. She has failure to thrive (height at the 5th percentile and a weight below the 5th percentile), severe congenital heart defects and heart failure (67%). The survival rate at age 1 day, 1 week, 1 month, and 1 year was 100%, 50%, 100%, and 92%, and 17%, respectively. Median survival time was 229.5 days (range, 22-2334). These data have a resemblance to the data about the patients with trisomy 18 in our hospital. (Kosho et al. 2006)
Predicting Obstructive Sleep Apnea in People with Down Syndrome. B. Skotko, S. McDonough, L. Voelz, D. Rosen, A. Ozonoff, E. Davidson, V. Alleradyce, N. Jayaratne, R. Bruun, N. Ching, G. Weintraub, L. Albers Prock, R. Becker, D. Gozal. 1) Medical Genetics, Massachusetts General Hospital, Boston, MA; 2) Boston Children’s Hospital, Boston, MA; 3) The University of Chicago, Chicago, IL; 4) David Geffen School of Medicine at University of California, Los Angeles, CA; 5) Harvard Medical School, Boston, MA.

Obstructive sleep apnea (OSA) in individuals with Down syndrome is associated with multiple morbidities: systemic and pulmonary hypertension, glucose intolerance, cardiovascular and cerebrovascular disease, and behavioral problems. The prevalence of OSA in this population is very high, with estimates ranging between 55-97%. Currently, an overnight polysomnogram (sleep study) is the gold-standard diagnostic test for patients with Down syndrome. Yet, this testing is cumbersome, poorly tolerated by these children, costly, and not widely available around the country. In this study, we looked to identify predictive factors for OSA in persons with Down syndrome. We enrolled ~100 subjects, ages 3-35 years, who already participated in the Down syndrome Program at Boston Children’s Hospital. For each patient, we collected subjective and objective measurements using validated parental survey instruments, standardized physical exams, lateral cephalograms, 3D-digital photogrammetry, and urine samples. Afterwards, all participants underwent standardized polysomnography at the Boston Children’s Hospital Sleep Laboratory where objective measurements were collected on OSA. We analyzed which combination of our assessment methods best predicted OSA, as ultimately determined by polysomnography. This will be the first time presenting the results of our data. Our final screening tool will hopefully allow physicians to avoid ordering a sleep study for those individuals with Down syndrome at lowest risk of OSA. Further, those patients with Down syndrome and clear predictors for OSA can proceed directly toward adenotonsillectomy, the current treatment.


Diamond-Blackfan anemia (DBA) is a rare congenital anemia due to arrested maturation of erythrocytes. Classic cases present as profound anemia in the first 1-2 years of life. Diagnosis is made with complete blood count demonstrating macrocytic anemia and reticulocytopenia in the setting of bone marrow deficient in erythroid progenitors. There may be associated physical malformations including thumb abnormalities, urogenital defects, craniofacial dysmorphism, low birth weight, growth and developmental delay. There is also a modest increase risk of acute myeloid leukemia, osteogenic sarcoma, and other malignancies. In the majority cases of DBA there is no family history and the causative gene mutations are de novo; however, in 10-25 percent of cases there is a positive family history with autosomal dominant inheritance and variable expressivity. Approximately 50 percent of DBA individuals possess mutations in genes coding for ribosomal proteins (DBA1-10); DBA4 due to mutations in the RPS17 gene at 15q25.2, accounts for less than 1 percent of cases. We report a 4 year old boy who initially presented with upper respiratory illness and low blood counts thought to be related to viral suppression. Symptoms (pallor, shortness of breath and leg pain) continued and he was referred to pediatric hematology where work-up for developmental delays revealed a de novo 46,XY, t(15:16) (q26.1;q13) karyotype. Although this reciprocal translocation appeared cytogenetically balanced, whole genome oligonucleotide array was performed (q26.1;q13) karyotype. Although this reciprocal translocation appeared cytogenetically balanced, whole genome oligonucleotide array was performed to rule out a cryptic imbalance of genetic material and revealed a 3.3 Mb de novo interstitial deletion extending from cytoband region 15q25.1 to 15q25.2. The deleted region contains more than 25 genes including the RPS17 gene associated with DBA4. Recurrent microdeletion of 15q25.2 has been described in association with a variable presentation that can include DBA; anomalies including congenital diaphragmatic hernia, VSD, and cryptorchidism;m; developmental and behavioral abnormalities; and short stature.

The Developmental Genome Anatomy Project (DGAP): Annotating the Human Genome from Balanced Chromosomal Rearrangements. K.E. Wong, I. Blumenthal, H. Brand, B. Currall, C. Hanscom, T. Hoyos, D. Lucente, Z. Ordulu, M.R. Stone, S. Pereira, V. Pillaiamamm, L.P. Yuan, J.F. Gusella, D.J. Harris, E.C. Liao, R.L. Maas, B.J. Quade, M.E. Taikowski, C.C. Morton, J.K. Seides, J. Davidson, B. Weintraub, 1) Department of Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 3) Departments of Psychiatry, Neurology, and Genetics, Harvard Medical School, Boston, MA; 4) Department of Plastic and Reconstructive Surgery, Center for Regenerative Medicine, Massachusetts General Hospital, Harvard Stem Cell Institute, Harvard Medical School, Boston, MA, USA; 5) Department of Pathology, Brigham and Women’s Hospital, Boston, MA; 6) Program in Medical and Population Genetics, Broad Institute, Boston, MA; 7) Division of Genetics, Boston Children’s Hospital, Harvard Medical School, Boston, MA, USA; 8) Division of Genetics, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA.

DGAP (www.DGAP.Harvard.edu) is a collaborative effort to identify genes critical in human development and disease pathogenesis through study of genes disrupted by balanced chromosomal rearrangements. Since 2001, 265 subjects have been enrolled, with recurring phenotypes that include craniofacial abnormalities (64%), symptoms of neurodevelopmental disorder (NDD) (70%) - 45%; comorbidity of both traits -, and other anomalies, including hearing loss (11%). Detection of breakpoints have migrated from FISH mapping and Southern blotting to next-gen sequencing, and 165 candidate genes and 21 non-coding RNAs have been elucidated. Sequencing of 99 genotypes from different individuals harboring a de novo balanced rearrangement has identified a rich landscape of chromosomal reorganization previously undetectable by cytogenetic methods. We detected complex chromosomal exchanges in ~20% of breakpoints at sequence resolution (previous estimates using cytogenetics predicted a rate of 2.8% complex exchanges), 90% of which contained microinversions, predominantly repaired through non-homologous end joining, and defined that highly complex localized chromosome ‘shattering’, or chromothripsis, resolves to largely balanced genomes from different individuals harboring congenital anomalies. To date, 26 animal models have been acquired or created for functional analysis of candidate genes. DGAP provides a unique opportunity to annotate the human genome from an ‘n of one’ perspective with exciting gene discoveries including a component of the BRAF-histone deacetylase complex as the critical intellectual disability gene in the Potocki-Shaffer syndrome (RPP21A), a histone/lysine demethylase/methyl-transferase in a profound debilitating phenotype (KDM6A), genes for traits such as height (HMGA2), a noncoding RNA in a severe developmental disorder (LINC00299), a chromdomain protein associated with autism spectrum disorder (CHD8), and single gene contributors to previously defined regions of genomic disorders (CDK6, MBDS, KIRREL3, EHM1, SATB2, and SNURF-SRPB). Although largely a basic research endeavor, DGAP methods have recently been implemented as clinical diagnostic tests, particularly for counseling dilemmas encountered with prenatal detection of de novo balanced rearrangements. With research results returned to physicians, DGAP offers a wealth of information on poorly annotated, clinically relevant areas of the genome and is actively seeking enrollments and collaborators.
Establishing a reference group for distal 18q-: clinical description and molecular basis. J. Cody1,2, M. Hashi1, B. Soileau2, P. Heard3, E. Carter4, C. Sebold5, L. O’Donnell2,2, B. Perry5, R. Straton5, D. Hale1, 1Dept. Pediatrics, Univ Texas Hlth Sci Ctr, San Antonio, TX; 2Dept. Psychiatry, Univ Texas Hlth Sci Ctr, San Antonio, TX; 3The Chromosome 18 Registry and Research Society, San Antonio, TX; 4Ear Medical Group, San Antonio, TX. Although constitutional chromosome abnormalities have been recognized since the 1960s, clinical characterization and development of treatment options have been hampered by their obvious genetic complexity and relative rarity. Additionally, deletions of 18q are particularly heterogeneous, with no two people having the same breakpoints. We identified sixteen individuals with deletions that, despite unique breakpoints, encompass the same set of genes within a 17.6 Mb region. This group represents the most genotypically similar group yet identified with distal 18q deletions. As the deletion is of average size when compared with other 18q deletions, this group can serve as a reference point for the clinical and molecular description of this condition. We performed a thorough medical record review as well as a series of clinical evaluations. Common functional findings included developmental delays, hypotonia, growth hormone deficiency, and hearing loss. Structural anomalies included foot anomalies, ear canal atresia/stenosis, and hypoplasias. The majority of individuals performed within the low normal range of cognitive ability but had more serious deficits in adaptive abilities. Of interest, the hemizygous region contains 38 known genes, 26 of which are sufficiently understood to tentatively determine dosage sensitivity. The data suggest that 20 are unlikely to cause an abnormal phenotype in the hemizygous state and five are likely to be dosage sensitive: TNX3, NETO1, ZNF407, TSHZ1, and ATP9B. ATP9B may be conditionally dosage sensitive. Not all distal 18q patients have the same clinical findings, and by comparison, this group provides an important advance in the molecular characterization of 18q deletions.


Terminal deletion of the long arm of chromosome 4 is a rare condition and the clinical findings may vary depending on the region that is missing. Developmental, craniofacial, digital, skeletal and cardiac involvements are the most common anomalies observed in these patients. Most of the common clinical findings which described for 4q terminal deletion syndrome are restricted to the 4q33---qter. A 7 year old male patient from non-consanguineous parents presented with developmental delay and multiple congenital anomalies. He had been operated for strabismus and hypothyroidism. At present, he has his first grade education. His height was 122 cm (25th centile), weight was 26 kg (50th centile), and head circumference was 53 cm (25th centile). He had mild motor delays, hypotonia, and hearing loss. Although general development is reportedly normal, were conceived through in vitro fertilization. Head circumference at age 4, that patient had relative macrocephaly (75th centile). The couple’s infertility of failure to achieve pregnancy was diagnosed by some Fragile X carriers; microarray analysis of both parents to assess the status of all CNVs is pending. This case further shows haploinsufficiency of FMR1 leads to altered dosage resulting in developmental delay. CMA is a sensitive method to detect copy number changes in FMR1.
2666F Microcephaly associated with duplication of chromosome 15q24. Y. Wang1, Y. Zhou2, D. Shrestha3, E. Carter4. 1) Dept Pathology, Univ South Alabama, 600 Clinic Dr, TRP IV, 278, Mobile, AL 36688; 2) Dept. of Biology, Baylor University, Waco, TX 76706. An 11-month-old African American female was born with microcephaly, and she has developmental delay. The physical examination shows an OFC of 43.5 cm (2nd percentile); height is 76 cm (90th percentile); and weight 11.9 kg (above the 95th percentile). The features previously noted for a child with craniofacial dysmorphism including narrow bitemporal diameter and sloping forehead with associated microcephaly. She has no associated anomalies or significant dysmorphisms, otherwise. A brain MRI revealed evidence of simplified gyral pattern with shallow sulci and a fewer gyri as well as thin corpus callosum. These findings are consistent with the diagnosis of microcephaly with simplified gyral pattern (MSGP). A detailed analysis of the constitutional chromosomal changes in the patient was conducted by microarray comparative genomic hybridization (array CGH), and linkage analysis using 6 short tandem repeat (STR) markers on chromosome 15. Array-CGH revealed microduplication with 1.265 Mb on chromosome 15q23-q24.1 (chr15:72023070-73288923 hg19 coordinates). Linkage analysis confirms this duplication. Human chromosomal region 15q24-26 harbors a high density of chromosome-specific duplisons. Consistent and recognizable clinical phenotypes of microduplication for distal 15q are minor craniofacial anomalies, congenital heart disease, mental retardation, and genital anomalies. Clinical phenotypes may vary in patients, depending on the size and location of duplication portion of chromosome 15q. Duplications in this region had been reported to be associated with neurodevelopmental disorders. However, microcephaly or brain malformation had not been reported in the 15q24 microduplication. This patient presents microcephaly and developmental delay caused by 15q24 microduplication.

2667W Elevation of Insulin-Like growth Factor Binding Protein 2 Level in Pallister-Killian Syndrome: Implications for the Postnatal Growth Retardation Phenotype. K. Izumi1,2, E. Kellogg3, M. Kaur4, A. Wilkens5, I. Krantz6. 1) Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Research Center for Epigenetic Disease, Institute of Molecular and Cellular Biosciences, The University of Tokyo, Tokyo, Japan. Pallister-Killian syndrome (PKS) is a multi-system developmental disorder caused by translocation 12p that exhibits tissue-limited mosaicism. The spectrum of clinical manifestations in PKS includes craniofacial dysmorphism, cleft palate, ophthalmologic, audiologic, cardiac, musculoskeletal, diaphragmatic, gastrointestinal, genitourinary, cutaneous anomalies, intellectual disability and seizures. Probands with PKS often demonstrate a unique growth profile consisting of macrosomia at birth with deceleration of growth postnatally. Since the insulin-like growth factor signaling (IGF) pathway plays a critical role in normal growth physiology, we hypothesized that dysregulation of IGF pathway is involved in the pathogenesis of PKS. Genome-wide expression array analysis performed on cultured skin fibroblasts from PKS probands revealed significantly elevated expression of insulin-like growth factor binding protein 2 (IGFBP2). The amount of IGFBP2 secreted from cultured skin fibroblast cell lines was measured. About 80% of PKS fibroblast cell lines secreted higher levels of IGFBP2 compared to the control fibroblasts, although the remaining 40% of PKS samples produced comparable level of IGFBP2 to that of control fibroblasts. Serum IGFBP2 level was measured in PKS probands, and in 40% of PKS probands, IGFBP2 level was elevated. No correlation was seen between mosaic ratio and the IGFBP2 level. PKS probands with elevated IGFBP2 manifested with severe postnatal growth retardation. IGFBPs are the family of related proteins that bind IGFs with high affinity, and IGFBPs are typically thought to attenuate IGF action, and mouse models of IGFBP2 overexpression demonstrate a postnatal growth attenuation phenotype. Hence, we suggest the possibility that elevated IGFBP2 levels might play a role in the growth retardation phenotype of PKS. Since IGFBPs functions in an endocrine/paracrine fashion, dysregulated IGF signaling pathway should manifest its effects not only on tetrasomic cells, but also on the chromosomally normal cells in PKS probands.

2668T Aetiology of Moderate Mental Retardation. A. KUMAR, M. VASHIST. DEPARTMENT OF GENETICS, M.D.UNIVERSITY, ROHTAK, INDIA. Mental retardation is a common disorder which imposes a large medical, psychological and social burden. It is grouped into four classes on the basis of IQ scores by ICD-10. Individuals with IQ (35-50) were placed in the moderate mental retardation (MMR) category. Two hundred moderately mentally challenged patients were studied from Pandit Bhagwat Dayal Sharma University of Health Sciences, Rohtak and other rehabilitation centers across Haryana. In our study the outdoor patients reaching the local medical care services were between age group from 1 month of child to 50 years. Age group showed peak values below 16 years and a sharp decline after 50 years of age. Nearly 80% excess of affected males in the present study strongly suggested contribution of non specific X linked mental retardation. Higher percentage of males can also be an indication of the community based bias towards a male child. There were 44% cases from lower socio economic classes like agriculture laborers, construction laborers and other daily wages workers. 72% cases of urban population with lower socio-economic status is an indication of some certain specific risk factors associated with moderate mental retardation. Poverty and nutritional deprivation are environmental conditions which also seem to be relevant predisposing risk factors. Various Non Government Organization and Social Welfare Departments must be intervened to produce more impactful programs in addition to existing schemes in the State of Haryana. The initiation of various health care programmes may limit occurrence of moderate mental retardation. Retrospective analysis revealed prenatal risk factors in 60% patients, neonatal in 40.35% and postnatal risk factors in 30.53% MR patients. Ninety four percent MR patients showed delayed developmental history. Malnutrition at prenatal stage was the main risk factor.

2669F A dyslexia case with de novo der(14)(Y;14)(q10;q10). E. Kirat, G. Güven, M. Seven, M. Özen, E. Yosunkaya, H. Ulucan. Medical Genetics, Istanbul University Cerrahpaşa Medical Faculty, Fatih, Istanbul, Turkey. Dyslexia is a neuropsychiatric syndrome which has historically been called as ‘congenital word blindness’ and contemporarily it is defined as developmental analogue of acquired selective impairment in reading and writing ability due to neuronal damage in certain regions of the brain. The etiology of dyslexia, whether it is a single trait or a cluster of subtypes which are likely to involve different subsets of genes, is an ongoing debate. Also in some of chromosomal abnormalities like XXY syndrome, dyslexia could be a symptom. XXY syndrome is a chromosomal aneuploidy with a predisposition to learning disability, tall stature, behavior disorder and long ear. Here we report a case that has a minor facial dysmorphism and dyslexia with an extra Y chromosome, translocated to the 14th chromosome with a karyotype 47,XY,Y. Here we report a case of a girl with XYY-like phenotype is under investigation.
2670W
Computer-aided facial recognition of individuals with Angelman Syndrome. L. Wolf1,2, W.H. Tan-3, L. Karna6, M. Shohat1-5, L.M. Bird1,5, 1) FDNA Ltd., Herzliya, Israel; 2) Tel Aviv University, Tel Aviv, Israel; 3) 3.NIH Rare Diseases Clinical Research Network–Angelman, Rett, & Prader-Willi Syndromes Consortium, USA; 4) Division of Genetics, Boston Children’s Hospital, Boston, MA, USA; 5) S.Schneider Children’s Medical Center of Israel, Rafael Recanati Genetics Institute, Rabin Medical Ctr and Felsenstein Medical Research Center, Petah Tikva, Israel; 6) Division of Genetics/Dysmorphology, Rady Children's Hospital San Diego, San Diego, CA, USA.

Angelman syndrome (AS) is a neurodevelopmental disorder characterized by severe developmental delay, intellectual disability, speech impairment and a unique behavior with an inappropriate happy demeanor. AS patients also commonly present unusual facies that may include deep set eyes, proptosis, a thin vermilion border of the lips, and broad flat nasion.

In this study we examined whether a computer-based dysmorphic analysis can help distinguish between AS patients and non-AS patients. We used a collection frontal facial images of genetically and clinically verified AS cases (n=210) and compared them to two separate control sets: (i) normal individuals (n=520); and (ii) individuals affected with other syndromes with dysmorphic facial features (n=808). The latter control set is added to verify that the system distinguishes not only between dysmorphic and non-dysmorphic populations, but can also identify the specific facial abnormality associated with AS.

A novel facial dysmorphology analysis system based on 2D photographs (FDNA®) was used. The system is fully automatic and starts by localizing hundreds of facial fiducial points and taking various measurements. The facial classification is based on the appearance of the entire facial image. A cross validation classifier is employed to discriminate Down syndrome and the healthy group by analyzing the selected features. The performance is evaluated using a method of automated detection of facial abnormalities that may be associated with phenotypic abnormalities. The availability of samples of monozygotic twins discordant for trisomy 21 has allowed us to study the effects of the trisomy in gene expression and genome organization in a unique model free of inter individual variability. When compared to the euploid twin, 374 peaks were enriched in the DNA from the fibroblasts of the twins, we found that the differential expression is organized in chromosomal domains with regions of upregulated or downregulated expression in the T21 twin. These domains correlated well with previously described replication domains and LADs. These domains could be responsible for tissue specific gene expression in the T21 twin. These domains have been used we hypothesized that mosaic chromosomal abnormalities are the differences in H3K4me3 levels and not the LADs topology are not altered in T21 cells, suggesting that the differences in expression between the twins are the result of changes in histone marks density or DNA methylation. In order to test the eventual effect of histone methylation on differential gene expression, ChIP seq for H3K4me3 was performed in the twins fibroblasts. HTS libraries for input and IP chromatin of both twins were sequenced in a Hiseq2500 36bp single read. 41 million reads per sample on average. Peaks were called using HOMER. 18,985 and 22,017 peaks were identified for the T21 and the euploid twin, respectively. These newly identified peaks are in common. These significantly enriched in the euchromatin (46%) are within or near genes related to embryonic development and morphogenesis according to DAVID. Notably, 3,788 peaks are unique to the euploid twin. They are significantly enriched in genes involved in proliferation, axogenesis, axon guidance and neuron morphogenesis. Furthermore, we estimated the fold change difference in H3K4me3 density between the twins for each gene. Remarkably, we found a strong correlation between the H3K4me3 density and gene expression for tissue specific gene expression patterns in the twins. More histone marks are currently being investigated to better understand the dynamics of gene expression dysregulation.

2672F
Tetraploid/diploid mosaicism as a potential cause of hypospadias. J.C. Giltay,1,2, A.J. Klijn,1,3, M. van Breugel,1,4, L. van der Veken1,2, R. Hochstenbach1,4, 1) Department of Genetic Medicine and Development, University of Geneva Medical School and University Hospitals of Geneva, Geneva, Switzerland; 2) National Center for Competence in Research Frontiers in Genetics, Geneva Switzerland; 3) Department of Genetic Medicine and Laboratories, University Hospitals of Geneva, Switzerland; 4) IGE3 institute of Genetics and Genomics of Geneva, Switzerland.

The use of samples from twins offers the opportunity to study epigenetic mechanisms that may be associated with phenotypic abnormalities. The availability of samples of monozygotic twins discordant for trisomy 21 has allowed us to study the effects of the trisomy in gene expression and genome organization in a unique model free of inter individual variability. When compared to the euploid twin, 374 peaks were enriched in the DNA from the fibroblasts of the twins, we found that the differential expression is organized in chromosomal domains with regions of upregulated or downregulated expression in the T21 twin. These domains correlated well with previously described replication domains and LADs. However, we demonstrated that LAD topology is not altered in T21 cells, suggesting that the differences in expression between the twins are the result of changes in histone marks density or DNA methylation. In order to test the eventual effect of histone methylation on differential gene expression, ChIP seq for H3K4me3 was performed in the twins fibroblasts. HTS libraries for input and IP chromatin of both twins were sequenced in a HiSeq2500 36bp single read. 41 million reads per sample on average. Peaks were called using HOMER. 18,985 and 22,017 peaks were identified for the T21 and the euploid twin, respectively. These newly identified peaks are in common. These significantly enriched in the euchromatin (46%) are within or near genes related to embryonic development and morphogenesis according to DAVID. Notably, 3,788 peaks are unique to the euploid twin. They are significantly enriched in genes involved in proliferation, axogenesis, axon guidance and neuron morphogenesis. Furthermore, we estimated the fold change difference in H3K4me3 density between the twins for each gene. Remarkably, we found a strong correlation between the H3K4me3 density and gene expression for tissue specific gene expression patterns in the twins. More histone marks are currently being investigated to better understand the dynamics of gene expression dysregulation.
of the literature revealed three reports of a chromosome 5;15 translocation involving the maternal chromosome 15 and is carried by her mother. Review c.819_822delCTGGinsGGTC (p.Asn273_Trp274delinsLysVal). Parental Sequencing of OCA2 revealed a previously reported hemizygous mutation: deletion of 15q11.2q13.1 that involved OCA2 (chr15:22,770,421-

pigment, fine curly yellow hair, blue eyes, and hypotonia. Her karyotype number of patients. We present here an African-American female, born at

These two conditions have been described together previously in a limited

declined visual acuity. This condition is due to mutations of OCA2 (previ-

albinism type 2 is diagnosed based on the clinical findings of skin and hair

of imprinted genes on the paternal 15q11.2-q13 region. Oculocutaneous

short stature, and characteristic facial features, is caused by the absence

feeding difficulties followed by later onset of hyperphagia and obesity with

ing needing assistance on unlevel ground or transitioning from carpet to
tile. 9/21 are partly self independent, being able to feed and dress them-

selves, but needing supervision and some assistance with personal hygiene,

and, at times, with walking on unlevel ground. 1/21 is fully independent. All

had variable degrees of developmental delays, with a moderate-severe
cognitive deficit. 18/21 enjoy a good health, whereas 1/21 has diabetes type
II, diagnosed at age 13. 1/21 has Raynaud disease, diagnosed at age 18;
and 1/21 had esophagitis at age 19. All but one are seizures free. All are

enrolled in a personalized rehabilitation program. Knowledge of the medical

and developmental aspects of adults with WHS will inform health supervi-
sion.

Prader-Willi syndrome and oculocutaneous albinism due to a 5;15 translocation and hemizygous OCA2 mutation. A.C.E. Hurst1, C.R. Hal-
deman-Englert2, T.H. Stamper2, M. Hanna3, M.J. Pettenucci2, P.P. Koty3, 1) Dept. of Pediatrics, Wake Forest School of Medicine , Winston-Salem, NC; 2) Department of Pediatrics, Section on Medical Genetics, Wake Forest School of Medicine, Winston-Salem, NC; 3) Wake Forest School of Medicine, Winston-Salem, NC.

Prader-Willi syndrome (PWS), characterized by infantile hypotonia and feeding difficulties followed by later onset of hyperphagia and obesity with delayed motor milestones, distinctive behavioral phenotype, hypogonadism, short stature, and characteristic facial features, is caused by the absence of imprinted genes on the paternal 15q11.2-q13 region. Oculocutaneous albinism type 2 is diagnosed based on the clinical findings of skin and hair hypopigmentation, nystagmus, and reduced iris and retinal pigment with decreased visual acuity. This condition is due to mutations of OCA2 (previously called the P gene), located within the same region of 15q11.2-12.

These two conditions have been described together previously in a limited number of patients. We present here an African-American female, born at 41 weeks gestation, who was noted at birth to have decreased cutaneous pigment, fine curly yellow hair, blue eyes, and hypotonia. Her karyotype revealed a 5;15 translocation: 45XX,der(5)/5(15)t(q53.5;q11.2). FISH analy-
sis for a PWS deletion was positive, and confirmed the deletion to be on the translocated chromosome 15. A cytogenetic microarray analysis clarified the loss of chromosome 15 material, showing a heterozygous 5.8-Mb deletion of 15q11.2q13.1 that involved OCA2 (chr15:22,770,421-

28,547,716; Build 37;hg19), with no gain of material from chromosome 5. Sequencing of OCA2 revealed a previously reported hemizygous mutation: c.819_822delCTGGinsGGTC (p.Asn273_Trp274delinsLysVal). Parental studies are pending, but it is likely that the translocation involves the paternal chromosome 15 and is de novo, and the hemizygous OCA2 mutation involves the maternal chromosome 15 and is carried by her mother. Review of the literature revealed three reports of a chromosome 5;15 translocation associated with PWS, but none of those patients had albinism. Therefore, our patient’s findings are a novel cause of the PWS plus albimism phenotype.

Comparing serum calcium tests in 22q11DS and other genetic condi-
tions: ionized vs. total calcium. E. Chow1,2, T. Leung1, M. Torsos1, C. Stefan1,2, 1) Clinical Genetics & Service, Centre for Addiction & Mental Health, Toronto, ON, Canada; 2) Department of Psychiatry, University of Toronto; 3) Clinical Laboratories and Diagnostics Services, Centre for Addiction & Mental Health, Toronto, ON, Canada.

Background: Hypocalcemia is common in 22q11.2 deletion syndrome (22q11DS) and can occur or recur throughout the life, making its detection an important part of routine monitoring. While the more expensive serum ionized calcium test is often recommended, no scientific research has been conducted on whether it has advantages over the cheaper and more easily performed total serum calcium test for patients with 22q11DS or other genetic conditions. This study aims to compare the detection of hypocalcemia by these two tests. Methods: A simultaneous set of serum calcium tests (ph-

normalized ionized calcium and albumin-corrected total calcium) was per-
formed 47 times in 26 individuals (9 male) with 22q11DS and 17 times in 16 individuals (9 male) mostly with other genetic conditions. All subjects were patients at the Clinical Genetics Service of CAMH in Toronto. Mean age at testing was younger for the 22q11DS group [24.9 (SD 6.6) years vs. 31.5 (SD 8.2) years, p=0.006] but the two groups did not differ in the propor-
tion of male subjects or subjects with psychosis. Measured hypocalcemia was defined as having one or both calcium test results below the lower limit of the reference range for the test. The number of times results from the two tests were discordant for measured hypocalcemia was recorded and compared. Results: Rate of measured hypocalcemia was higher in the 22q11DS group (36.2% vs. 11.8%), but at a trend level (p=0.07). For the 22q11DS group, test results were discordant in 9 of the 17 instances of measured hypocalcemia: 7/9 in the ionized calcium test and 2/9 in the total calcium test. 22q11DS subjects had neuromuscular symptoms in 2 out the 7 times when only the ionized calcium was abnormal, but none when only the total calcium was abnormal. On the other hand, both instances of hypocalcemia in the other group were found only in one test, and both times in the ionized calcium test. Although neither of these two subjects was symptomatic, the one with Rett Syndrome had a significantly elevated PTH. Conclusions: The increased instances of hypocalcemia detected in the 22q11DS group were normalized by the ionized calcium test and two true hypo-
calcemia as subjects were more likely to have symptoms of hypocalcemia and/or an elevated PTH. Thus the pH-corrected ionized calcium test was a more sensitive and superior test in detecting hypocalcemia in 22q11DS and other genetic conditions than the albumin-corrected total calcium test.

Scoliosis and Vertebral Anomalies: Additional Abnormal Phenotypes Associated with Chromosome 16p11.2 Rearrangement. H. Al-Kateb1, G. Khan1, I. Filges2, S. Kulkarni3, M. Shirani1. 1) Pathology and Immunology, Washington University School of Medicine, St. Louis, MO; 2) University of British Columbia, Vancouver, BC.

16p11.2 rearrangements are estimated to occur at a frequency of approxi-
mately 0.6% of all samples tested clinically and have been identified as a major cause of autism spectrum disorders (ASD), developmental delay, behavioral abnormalities, and seizures. Careful examination of patients with these rearrangements revealed association with abnormal head size, obe-
sity, dysmorphism and congenital abnormalities. In this report, we extend this list of phenotypic abnormalities to include scoliosis and vertebral anomalies through detailed characterization of phenotypic and radiological data of seven new patients, 5 with 16p11.2 deletion and 2 with duplication corre-
sponding to chromosomal coordinates: chr16: 29,528,190-30,107,184 (hg19) with a minimal size of 579 kb. We discuss the phenotypic and radiological findings in our patients and review 5 previously reported patients with 16p11.2 rearrangement and similar skeletal abnormalities. The results show a male to female (M:F) ratio of 2:1, strikingly different from the known 1:10 ratio in the general population for idiopathic scoliosis. Interestingly, we found that males have more severe spinal deformities than females. We discuss the implications of these findings on the diagnosis, surveillance and genetic counseling of patients with 16p11.2 rearrangement.
A novel interstitial microdeletion in 2q37 refines critical region and candidate genes for microcephaly, myelination and developmental delay expressed in human neural progenitors. J. Imotola1, D. Khurana2, A. Legido1, K. Carvalho1; 1) Brigham and Women’s Hospital Harvard Medical School; 2) Section of Pediatric Neurology, St Christopher Hospital for Children’s, Drexel School of Medicine, Philadelphia.

2q37 microdeletion syndrome is characterized by bone, cardiovascular alterations, neurodevelopmental delay, microcephaly and seizures. This syndrome is associated with loss of genetic material approximately 100 genes in the 2q37 band. However, the genes associated with neurodevelopmental phenotype in this syndrome are still unknown. A novel 496 kb deletion was discovered by whole genome array CGH in a patient who fulfilled the criteria for 2q37 microdeletion syndrome with additional features of hand wringing, toe walking and seizures. The abnormal segment contains 10 genes based on UCSC 2006 hg18, including SEPT2, FARP2, HDLPB, STK25. Pathway analysis of these genes revealed links to neural development, myelination, cilogenesis and interaction with genes associated with neural function, gene expression analysis revealed that these genes are highly expressed in cortical plate, human neural progenitors are in vivo and in vitro. Our report narrows the genomic region for 2q37 microcephaly, myelination and neurodevelopmental delay to 10 candidates genes and suggests that this segment may represent an important locus for human forebrain development. Haploinsufficiency of this region may lead to human microcephaly, seizures and neurodevelopmental delay.

2681F Deep White Matter Brain Abnormalities in a Patient with Chromosome 15q11-q13 Deletion and Angelman Syndrome Phenotype. N. Sekhri3, S. Scharer3, D. Khurana1,2, E. Frengen1,2, J.R. Helle1, A. Legido1, S. Axelsson1,2, T. Barøy1,2, B. Delle-Chiaie1,2, E. Deacon1,2, J. Imitola1, V. van den Veyver1,2, J. McPherson1,2,3,1) Dept. of Pediatrics, Genetics Center, Children’s Hospital of Wisconsin, Milwaukee, WI; 2) Advanced Genomics Laboratory, Medical College of Wisconsin, Milwaukee, WI.

While Angelman syndrome (AS) is a neurodevelopmental disorder with a known genetic basis that is characterized by severe intellectual disability, lack of speech, ataxia, seizures, and specific behavioral patterns, leukodystrophy is not a typical finding. Subtle changes to white matter pathways detected by DTI (diffusion tensor imaging) have been described by Peters et al., 2011. We describe a 22-month old male patient with more overt leukodystrophy (detected on standard 1.5T brain MRI), childhood onset complex partial seizures and borderline microcephaly. MRI showed symmetric long TR hyperintensity in the periventricular and deep white matter involving the fronto-temporal and parieto-occipital white matter without evidence of abnormal enhancement. Diagnostic work-up initially focused on the white matter abnormalities and seizures; Angelman syndrome was not considered high in the differential diagnosis. Chromosomal microarray analysis was included in the work-up, but not methylation studies. Subsequently, the patient was found to have a 4.9 megabase deletion on microarray analysis (Affymetrix 8.0 oligo/ SNP array) at chromosome 15q11.2-q13.1 encompassing the Prader-Willi syndrome (PWS)/AS) region, while the remainder of genetic tests were negative. After review, the patient’s behavioral pattern and clinical presentation was felt to be consistent with a diagnosis of AS. This case highlights a potential emerging problem in the approach to confirming a suspected genetic diagnosis. The availability of clinical exome/gene expression analysis and sequencing of disease-targeted gene panels based on Next-Generation Sequencing (NGS) technologies still leave gaps in the detection of common disorders due to methylation defects, triple repeat expansion or small CNVs. Clinicians should keep these limitations in mind when evaluating patients with varied genetic disorders.
2682W
Deletion 2q37: Cognitive-Behavioral Profiles, Developmental Trajectories, and IQ Related to Deletion Size. G.S. Fisch1, S.T. South2,3, A. Rutherford3, R. Falt1, J. Carey2,1, New York University Dept. Epidemiology & Health Promotion New York, NY; 2) University of Utah Dept. Pediatrics Salt Lake City, UT; 3) Institute for Clinical and Experimental Pathology ARUP Laboratories Salt Lake City, UT; 4) Medical Genetics Institute Cedars-Sinai Medical Center Los Angeles, CA.

2q37 deletion syndrome is typically described as encompassing, among other clinical features, mild to moderate intellectual disability, behavioral problems, short stature, hypotonia, characteristic craniofacial anomalies, and a high proportion of affected individuals with autism. Deletions involving 2q37 occur most often as a de novo mutation although instances of inversions and unbalanced translocations have been observed. We report on 8 cases with deletion of 2q37, the largest single multicenter study to date of individuals with this genetic anomaly. We examined clinical features, cognitive and behavioral abilities from 5 females and 3 males, previously diagnosed cytogenetically with del2q37, and ranging in age from 4 - 18 years at the time of their first cognitive-behavioral assessment. Each participant was administered a comprehensive neurocognitive battery consisting of the Stanford-Binet (4th Ed; SBFBE), Vineland Adaptive Behavior Scales (VABS), Conners Parent Rating Scale [CPRS], Child Behavior Checklist [CBCL], and the Child Autism Rating Scale [CARS]. Of the 8 participants, 3 lacked expressive language and could not be assessed by the SBFBE. Intellectual disability for the 5 tested with the SBFBE ranged from severe to mild [IQ Range: 36 - 59], with relative strength in Quantitative Reasoning. Adaptive behavior composite scores from the VABS were much below adequate levels, ranging from a floor value [19] to 55. Based on the CPRS, 7/8 [91%] could receive a diagnosis consistent with ADHD. Scores on the CARS ranged from 22 [non-autistic] to 56 [extremely autistic]; 5/8 [63%] children received scores in the autism range. Deletion size based on microarray analysis using Cytoscan HD or Affymatrix 6.0 was determined for 4 subjects from whom DNA was obtained. Curiously, participants with the largest deletions, 10.1 Mb and 9.5 Mb, achieved the highest IQ and DQ scores; whereas, those with the smallest deletions, 7.9 Mb and 6.6 Mb, attained the lowest IQ and DQ scores. Previously, the CENT2 gene [now known as AGAP1] was associated with a deletion phenotype severity is discussed further.

2683T
Identification of 22q11.2 deletion in patients from adult congenital heart disease clinic - a missed burden in the transition care in Hong Kong. B. Chung,1, P. Chow,2,3 1) New York University Dept. Epidemiology & Health Promotion New York, NY; 2) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Toronto, Ontario, Canada; 3) University of Toronto, Toronto, Ontario, Canada.

We describe two unrelated cases of de novo microdeletions of chromosome 22q11.2, one at band q22.1 and the other at band q11.2, in adult patients with conotruncal defects and with distinctive and recognizable features. One patient displayed hearing loss, a history of intrauterine growth retardation, failure to thrive, developmental delay, and congenital heart defect. At her last assessment at 4 years of age, the patient met criteria for autism spectrum disorder, had significant global intellectual disability and was tube fed. Also she had hearing and visual impairment. Microarray analysis (44,000 oligos platform) revealed an estimated 0.444 Mb deletion of chromosome 16q at band q22.1, which included 25 RefSeq genes. In both patients, the deletion comprised 4 OMIM genes involved in cardiovascular malformations consisting predominantly of conotruncal heart defects, palatal abnormalities, characteristic facial dysmorphic features, learning difficulties, mild to moderate immune deficiency, hypoparathyroidism, among other. About 90% of patients with the 22q11.2 microdeletion have a common ~3 Mb deletion, whereas 7% have the patients have a smaller 1.5 Mb deletion and the remaining 3% have other atypical deletions, nested, overlapping or adjacent to the typically deleted region (TD1). However, to date, only a limited number of atypical deletions in 22q11.2 have been reported and only a few are located distally to TD1. In this paper we report a male 47 years old patient who had been studied for primary immunodeficiency with recurrent Herpes Zoster infection which onset 9 years ago. The patient had a history of delayed developmental milestones, mild learning difficulty at school, hypothyroidism and unilateral sensorineural hearing loss. At the first consultation the patient display slight facial dysmorphism, including hooing of the lower lids, long pointed nose, velopharyngeal voice, left ocular Herpes Zoster infection with fourth nerve palsy and mild cognitive deficiency. He had a 46, XY normal karyotype. Immunological profile reported remarkable decrease of immunoglobulin and CD3, CD4 and CD8 deficiency. His IQ total score was 72. These findings suggested a DiGeorge Syndrome, therefore was performed an array-based comparative genomic hybridization (aCGH). It reported a ~1.8 Mb deletion within chromosome band 22q11.21-22q11.23 that matched with distal 22q11.2 microdeletion (OMIM: 192431). Further, the microdeletion was confirmed by multiplex ligation-dependent probe analysis (MLPA) (probes M155_B1 DiGeorge). Here we report one of the few atypical microdeletions in the distal 22q11.2 deletion region with a barely phenotype and an important cellular immunodeficiency leading to recurrent Herpes Zoster infection that has not been described previously. The phenotypic spectrum of 22q11.2 deletion depends on multiple dosage-sensitive genes located in this region. Non-overlapping atypical deletions have showed significantly overlapping phenotypes, suggesting either, a positional gene effect in 22q11.2, two involvement of several candidate genes or a common developmental pathway.
Array comparative genomic hybridization (array CGH) is widely used as a first-tier clinical diagnostic test in unexplained developmental delay/intellectual disability (DD/ID) and/or congenital anomalies cases. We have tested by array-CGH (Agilent 60 K) a cohort of 513 patients affected by DD/ID associated with congenital malformations and/or dysmorphisms. The detected CNVs, unlisted in the Database of Genomic Variants, were assigned to one of the following four groups: 1) known microdeletion/microduplication syndromes or CNVs encompassing a disease-causing gene: 46/513 patients (8.9%), including del 22q11.21 (n=6), del/dup 15q13.3 (n=2+2), del/dup 16p11.2 (n=3+4), del 17p12.1.31 (n=2), del 1 q21.1 (n=1), del 3q29 (n=1), del 7q11.23 (Williams-Beuren syndrome, n=2), dup 1SD1 gene, del CREBBP gene (Rubinstein-Taybi syndrome), del NF2 gene, del PITX2 (Pieger syndrome), del ANKR/D1 gene (Koolen de Vries syndrome); 2) deletion/duplications extending more than 7 Mb (range 7.5-29 Mb): 23/513 patients (4.5%), including 2 cases of mosaicism (>30%). The remaining CNVs involved chromosomal regions of challenging study by standard karyotype. Five of 23 cases derived from a parental balanced translocation; 3) likely pathogenic deletion/duplications (150 K-1.4 Mb): 15/513 cases (2.9%), resulting de novo in 5/15 and inherited in 10/15. Interestingly we detected an atypical deletion of 1.3 Mb in 3q29 (encompassing 13 known genes), a deletion of 500 Kb in 5q12.3 (5 known genes) and a deletion of 1.4 Mb in 16p11.2 (14 known genes). 4) CNVs of unclear relevance (110-900 kb): 52/513 (10.1%). These results restate the effectiveness of array-CGH analysis as first-step test in patients presenting with complex phenotypes, not only to reach a clinical diagnosis but also to identify new candidate genes for DD/ID.

Microarray Comparative Genomic Hybridization analysis (array-CGH) could be a novel contiguous gene syndrome with renal impairment. N. Morisada 1, A. Kaito 2, H. Loyola 1. 1) Pediatrics, Toho University Ohashi Hospital, Tokyo, Japan; 2) Pediatrics, Toho University Ohashi Hospital, Tokyo, Japan.

Microdeletion of 16q is a rare chromosomal abnormality. Nonsense mutation of SALL1 at 16q12.1 causes Townes-Brock syndrome (TBS, OMIM #107480) with branchiogenic anomalies, imperforate anus, and mild renal impairment. The patients with SALL1 heterozygous entire deletion develop milder clinical symptoms. Renal impairment in the patients with SALL1 heterozygous entire deletion has not been reported. Homozygous mutation of RPRGIP1 at 16q12.2 cause Joubert syndrome (OMIM #611560) with nephronophthisis. There are few case reports with the deletion of both genes. Moreover, these several reported cases did not develop renal impairment. Case reports. Case 1: A 15-year-old Japanese boy had hearing loss, external ear malformation, mild developmental impairment and multiple congenital anomalies. Case 2: A 21-year-old Japanese woman had multiple abnormalities such as bilateral renal hypoplasia leading to end stage renal failure, severe developmental delay, imperforate anus, congenital heart defect, and low set ears. He was diagnosed as TBS. Fluorescence in situ hybridization (FISH) analysis revealed entire gene deletion of SALL1. A subsequent aCGH analysis revealed 6Mb deletion in 16q including SALL1 and RPRGIP1.

Discussion. Most of 16q heterozygous microdeletion cases with only one gene either SALL1 or RPRGIP1 have no renal impairment. However, our cases with both heterozygous gene deletions had severe to moderate renal impairment. Therefore, we suggest that there may be some interactions between SALL1 and RPRGIP1 gene, and our cases may be a new contiguous gene syndrome.
Fetal Alcohol Syndrome and Pitt-Hopkins Syndrome in four maternal half siblings, A. Asamoah, K.E. Jackson, Y. Senturia, K. Goodin, G.C. Gowans, K. Platky, J.H. Hersh. 1) Dept Pediatrics, Univ Louisville, Louisville, KY; 2) Dept Pediatrics, Carolinas Medical Ctr, Charlotte, NC.

Fetal alcohol syndrome (FAS) is identified by the presence of growth impairment, characteristic facial features and central nervous system abnormalities in the background of prenatal alcohol exposure. Fetal Alcohol Spectrum Disorders (FASD) is an umbrella term for the group of conditions arising from prenatal alcohol exposure that results in physical, intellectual, developmental and behavioral disabilities in the affected individual. Disorders on the spectrum include Alcohol-Related Neurodevelopmental Disorder (ARND), Partial FAS (PFAS) and Alcohol-Related Birth Defects (ARBD). Pitt-Hopkins syndrome is caused by haploinsufficiency of TC4 gene either due to a point mutation or a deletion of the chromosomal region 18q21.2. Affected individuals have distinctive facial features, developmental delay/ intellectual disability, and hyperventilation and/or breath-holding episodes while awake. We present four maternal half siblings diagnosed with FAS based on maternal history of alcohol abuse during pregnancy and physical examination who on chromosomal microarray analysis were found to have 18q21.1 deletion that includes the TC4 gene. The finding of TC4 deletion in these 4 maternal half-siblings suggests their mother has a germline deletion or is mosaic for the deletion herself. These findings suggest considering chromosomal microarray analysis in patients with FASD to determine if there are no concomitant chromosomal abnormalities contributing to the phenotype.

Normal appetite and BMI in a 9 year old girl with haploinsufficiency of SIM1 due to a 2.2 MB deletion at 6q16.2-q16.3. G.A. Bellus, K. Zega. Pediatrics, University of Colorado, Aurora, CO., USA.

Many patients with interstitial deletions involving 6q16 are reported to have phenotypic features that resemble Prader-Willi syndrome with obesity, short hands and feet, hypotonia and developmental delays. Several genotype-phenotype studies have implicated haploinsufficiency of SIM1 (homologue of Drosophila single minded /OMIM 603128) as a candidate gene for obesity and developmental delays in these individuals. In Drosophila, sim is a transcription factor that is considered to be a master regulator of neurogenesis. In the mouse, Sim1 is required for the development of the paraventricular nucleus of the hypothalamus and Sim1 heterozygous knock-out mice exhibit hyperphagia, early onset obesity and increased linear growth. On the other hand, individuals with obesity and developmental delays have been reported with deletions at 6q14.1-q15 that do not involve SIM1. This suggests that SIM1 haploinsufficiency may not be necessary to cause obesity and development delays in some individuals with deletions in this region. We report a 9 year old girl with a history of hypotonia and mild intellectual disabilities with a normal appetite and BMI who was found to have a 2.2 MB deletion at 6q16.2-q16.3 involving SIM1. Together these results suggest that haploinsufficiency of SIM1 may be neither necessary nor sufficient to cause hyperphagia and obesity in individuals with 6q deletions.

Lack of Nablus mask-like facial syndrome phenotype in a patient with a de novo microdeletion of chromosome 8q21.2q22.1. D. Cherukuri, B. Crandall, S. Kantarci, Department of Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 2) Departments of Pediatrics and Psychiatry, UCLA.

Nablus mask-like facial syndrome (NMLFS) is associated with deletions of 8q22.1 chromosome region. The clinical manifestations of this microdeletion syndrome include tight appearing glistening facial skin, blepharophimosis, telecanthus, sparse arched eyebrows, flat and broad nose, long philtrum, distinctive ears, upturned frontal hairline, short and broad neck, and developmental delay. Happy disposition is a typical behavioral trait of NMLFS [MIM: 608156]. Here, we report a 2-year old girl with a history of developmental delay and dysmorphic features. This girl was born post-term to a healthy nonsmoking couple of European descent. The delivery was vaginal and uncomplicated. At birth, the birth weight was 7 pounds 8 ounces, and there were no particular concerns postnatally and she was discharged in 2 days. She had delay in sitting at 9 months of age and sat alone at about 13-14 months period. Head circumference was 10 to 25th percentile, weight was 75th to 90th percentile, and height was 25th percentile at 22 month. Physical examination revealed lower limb hypotonia, trigonocephaly and dysmorphic facial features, including hypertelorism, depressed broad nasal bridge, low lying columella, mild prognathism, deep philtrum, wide mouth, thin upper vermilion border with prominent cupid’s bow, and broad and short neck. Her ears were normal. Currently, she had speech delays and distinctive happy and social personality. SNP-chromosomal microarray testing (SNP-CMA) with Affymetrix CytoScan HD array revealed a de novo 10.9 Mb deletion of chromosome 8q21.2q22.1 ranging from chromosome position from 86,845,949 to 97,724,872 (hg19). This deletion interval includes 51 RefSeq genes and overlaps with the reported NMLFS critical region (93.98-96.22 Mb (hg19)). Our patient does not present with the characteristic facial features of this syndrome. However, her happy and social personality is a common behavioral trait observed among the NMLFS patients. Similar to our findings, there are a few reported patients, without the NMLFS clinical features, with overlapping microdeletions with the NMLFS critical region. In conclusion, our report supports that deletions of 8q22.1 region result in variable phenotypes with and without the NMLFS syndrome.
2694W
Case report of a 17q21.31 microdeletion associated with EFTUD2 mandibulofacial dysostosis with microcephaly identified by comparative genomic hybridization. S.K. Gandomi1, D.M. Reeves2, M. Parra3, C.L. Gau4, V. Yap2. 1) Ambry Genetics, 15 Argonaut, Aliso Viejo, CA 92656; 2) University of Arkansas for Medical Sciences, Arkansas Children’s Hospital, 1 Children’s Way, Slot 512-5, Little Rock, AR 72202; 3) University of Arkansas for Medical Sciences, Department of Neonatal-Perinatal Medicine, 4301 West Markham, Slot 512-5B, Little Rock, AR 72205.

Mandibulofacial dysostosis with microcephaly (MFDM) is a rare, sporadic malformation syndrome manifesting with severe craniofacial abnormalities, microcephaly, developmental delay, and additional dysmorphic features. Although most cases of clinically diagnosed MFDM remain genetically unexplained, recent sequencing studies have linked this condition to heterozygous EFTUD2 mutations in 15 probands in the literature. In this case report, we present a previously undescribed dizygotic female twin proband (Twin A) born at 36 weeks gestation with severe microcephaly, microtia, cleft palate, severe retroglossa, oral and pharyngeal dysphagia, bilateral proximal radioulnar synostosis, 11 thoracic ribs, abnormal MRI findings, high-pitched cry due to unilateral vocal cord paralysis, and additional dysmorphic features. Newborn screening and a series of additional biochemical investigations were diagnostically negative. The proband’s twin sister (Twin B) was born healthy and shows no phenotypic similarities. Family history is unremarkable for any known genetic syndromes, and the twins’ parents are both reportedly in good health. Array comparative genomic hybridization (aCGH)+SNP analysis was performed on Twin A to assess for chromosome rearrangements and regions of homozygosity. Results of this assay identified a small de novo pathogenic deletion on chromosome 17q21.31, encompassing the EFTUD2 gene. The deleted region also included 13 million bases considered unlikely to be responsible for the proband’s phenotype. No regions of homozygosity were identified in the 400K array, which would also confirm a non-consanguineous family history. Of the total 15 reported FDM-associated EFTUD2 mutations described to date, all alterations resulted in genetic haploinsufficiency, consistent with our proband’s microdeletion pathomechanism. In addition, our proband’s phenotypic features both overlap and expand on the clinical features of previously reported probands, as well as new ones. Lastly, we present a series of additional biochemical investigations and reporting of other individuals with EFTUD2 mutations and clinical MFDM to better delineate genotype-phenotype correlations for more accurate diagnosis of this complex condition.

2695T
Nablus mask-like facial syndrome: 3 additional cases add support that del 9q22.1 is necessary but not sufficient to cause the classic phenotype. S.S. Janam1, H. Dukzale1, N. Dukzale1, C. Zhang2, F.A. High1, L. Kaban1, S. Bhattacharyya1, J.M. Stoler2, A.E. Lin3. 1) Harvard Medical School Genetics Training Program, Boston, MA; 2) Dept of Medical Genetics, Osmangazi University School of Medicine, Eskisehir, Turkey; 3) Oral Maxillo-facial Surgery, Massachusetts General Hospital, Boston, MA.

Conclusion: Deletion of 9q22.1 has been reported in patients (pts) with Nablus mask-like facial syndrome (NMLFS) with its striking craniofacial and dermatologic anomalies. 9 pts reported by Allanson et al. [AJMG 2012] with overlapping deletions, but without the complete NMLFS phenotype suggested that del 9q22.1 is necessary but not sufficient. We report 3 additional pts with del 9q22.1, 2 of whom had NMLFS; one is undergoing complex oral maxillofacial surgery (OMFS). CASES: Pt 1: 14 yo white male with jaw ankylosis, cleft soft palate, blepharophimosis, small ears, conductive hearing loss and shiny tight facial skin. NMLFS was suspected clinically and a 2.7 Mb del 9q22.1 was confirmed by chromosome microarray (CMA). Procedures prior to diagnosis included distraction osteogenesis, tracheostomy and cleft palate repair. Surgery after diagnosis included curvilinear osteogenesis and cleft lip and palate, bilateral cleft palate repair, bilateral microtia, and vertigo surgery. Pt 2: 20 yo male from Turkey with microcephaly, shiny tight facial skin, and epidermolysis bullosa acquisita. NMLFS was suspected clinically and a 2.7 Mb del 9q22.1 was confirmed by CMA. Pt 3: 56 yo male from Iran with severe craniofacial abnormalities, microcephaly, facial and airway anomalies. NMLFS was suspected clinically and a 2.7 Mb del 9q22.1 was confirmed by CMA.

Conclusions: Deletion of 9q22.1 is necessary but not sufficient to cause the classic NMLFS phenotype. Complex OMFS can be achieved with intensive multidisciplinary care, especially techniques for airway management (micrognathia, restricted neck movement, small mouth); tracheostomy is done as a last resort.

2696F
In the shadow of MEF2C: Genotype-phenotype correlation for 5q14.3q21 deletions. J.A. Rosenfeld1, K. Stoate4, A. Asamoah2, R.R. Lebo1, J.A. Faschin1, L. Russell3, J.W. Ellison1, L.A. Schinzel2, J. Sailor2, Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 2) University of Louisville, Louisville, KY; 3) SUNY Upstate, Syracuse, NY; 4) Pontificia Universidade Catolica do Parana, Curitiba, PR, Brazil; 5) Department of Medical Genetics, McGill University Health Centre, Montreal, Quebec, Canada; 6) University of Minnesota, Minneapolis, MN.

High resolution microarray analysis for genomic deletions and duplications has continued to provide ongoing delineation of genomic syndromes and generation of novel hypotheses regarding gene function in human development. We report further delineation of the phenotypic consequences of genomic deletions within 5q14.3q21.1. Microarray-based comparative genomic hybridization was performed on a sample from a 5.5-year-old male with developmental delay, severe expressive speech delay, visual impairment, bilateral incomplete coloboma of the iris, bilateral sensory hearing deficits, and seizures. The results identified four small interstitial deletions within 5q14.3q21.1, distal to MEF2C and totaling 2.1 Mb. These findings suggest that genes distal to MEF2C may also cause epilepsy and neurodevelopmental impairment. A review of the literature showed that some of the features in our patient have been occasionally reported with deletions extending to 15q21, including iris colobomas, hearing loss, and urogenital anomalies. To further delineate phenotypes associated with deletions in this genomic region we conducted a survey of patients in our database of clinical microarray cases and identified additional incidences of these features, though they were not fully penetrant. A comparison of our patients to those previously reported supports possible roles for developmentally important genes in telencephalic phenotypes, specifically NR2F1, with roles in neural, optic, and otic development, and RGMB, with roles in BMP/2/4 signaling. In particular, NR2F1 haploinsufficiency is likely associated with hearing loss, optic nerve atrophy, and hypoplastic corpus callosum observed in our proband. This report represents the importance of ongoing delineation of genomic syndromes through case identification and comparison to past cases through a robust clinical database.
2697W
Overlapping Phenotype of Silver-Russell-like and 14q32 Microdeletion Syndromes in a Child with Submicroscopic 11p15.5 Duplication and 14q32 Deletion. H.J. Mrozikowski1, D.B. Lowenstein2, H. Abdel-Hamid3, D.N. Safer1, A. Rajkovic1,2, S.A. Yatsenko1,2 1) Department of Obstetrics, Gynecology and Reproductive Sciences, Magee-Womens Research Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA; 2) Department of Perinatology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 3) Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA.

2699F
Bilateral cleft lip and bilateral thumb polydactyly with triphalangeal component in a patient carrying two de novo deletions on chromosome 4q32 and 4q34 involving PDGFC, GRIA2 and FBXO8 genes. A. Brusco1,2, A. Calcì1, G. Gai1, E. Di Gregorio1,2, F. Talarico1, V.G. Naretto1, N. Migone1,2, E. Pepe1, E. Grosso1,2 1) Dept. Medical Sciences, Univ Torino, Torino, TO, Italy; 2) A.O. Città della Salute e della Scienza, S.C.d.U. Medical Genetics, Torino, Italy.

We report a newborn male with bilateral clefts of the primary palate, a duplicated bilateral triphalangeal thumb, and a patent foramen ovale. At 4 yr., he presented moderate psychomotor developmental delay with normal brain MRI. The association of clefts of the lip/palate (CLP) and triphalangeal thumbs, per se an extremely rare finding, has never been reported so far. In our case, the array-CGH analysis revealed two de novo deletions (~1.2 Mb and ~400 Kb) on the long arm of chromosome 4, containing four genes: platelet-derived growth factor C (PDGFC), glycine receptor beta subunit (GLRB), glutamate receptor ionotropic AMPA 2 (GRIA2), and F-box protein 8 gene (FBXO8). PDGFC codes for a mesenchymal cell growth factor already associated with clefts of the lip. PDGf-/- mice have skeletal anomalies, and facial schisis resembling human cleft lip palate; GRIA2 codes for a ligand-activated cation channel that mediates the fast component of postsynaptic excitatory currents in neurons, and may be linked to cognitive dysfunction. FBXO8, a gene of unknown function, is member of the F-box gene family, among which SHFM3 is mutated in human split-hand foot malformations type 3. The presence of overlapping deletions in patients who do not share the same phenotype of our case suggest an incomplete penetrance, with a possible effect of modifier genetic factors.

2700W

Pure duplications of chromosome 20p have been reported rarely in the medical literature. We presented a family with a duplication of chromosome 20p13-p12.1 of 9.2 Mb in size. The proband was initially referred in childhood for a submucous cleft palate with bident uvula and unilateral high-grade myopia. She was later found to have a didelphic uterus, bilateral renal cysts, and cervical and thoracic vertebral anomalies. Her mother has a bicornuate uterus, bilateral high-grade myopia, and cervical anomalies. Physical features are shared with a maternal uncle, who has a VSD, and the maternal grandmother. The mouths are small, the columella are low hanging and the nasolabial creases are deep. There is brachydactyly and tapering of the digits. The voice quality is marked by hyponasality. Family members have pursued post-secondary education with success. The proband has significant anxiety, and this is seen to a lesser extent in her mother. Duplications of 20p are reported to be associated with cognitive impairments, but the regions involved are larger than what we describe. Vertebral anomalies are common in trisomies of this region and may reflect dosage effects of the JAG1 gene. The MKKS gene is known to cause a recessive phenotype due to loss of function mutations. However, this family demonstrates features of Bardet-Biedl syndrome (uterine anomalies, high palates, renal cysts, and anxiety), despite having an increase in copy number. They do not have polydactyly or retnitis pigmentosa, although ocular problems are present with extremely high myopia. In summary, large 20p duplications are associated with a number of anomalies, but a normal cognitive outcome is possible.
2701T
Inherited Yq112ter Deletion Associated with Congenital Cataracts, Microphthalmia and Autistic Spectrum Disorder in 3 Brothers. P. Biloum1, A. Delahaye2,2, B. Benzécri2,3, E. Pipiras2,2, 1 Gen Med, CHU Paris-Nord, Hospital Jean Verdier, BONDY, France; 2 Embryo-Cyto genetics, CHU Paris-Nord, Hospital Jean Verdier, Bondy, France; 3 INSERM 676, Robert Debre Hospital, Paris, France.

Purpose: We screened a cohort of 65 patients with syndromal ocular disorder for CNV by microarray analysis after informed consent. We identified a family of male twins and an older brother with congenital cataract and possibly secondary microphthalmia associated with autistic spectrum disorder with cognitive delay, absent speech, introverted affect and severe communication disorder respectively aged 29 and 30 years old and a novel CNV. Methods We used an Illumina microarray platform with Human Hap 300, Human Cyto SNP 12, with 300k whole genome markers upon DNA extracted from lymphoblastoid cell lines. Illumina Software with Genome Studio 2010.3, CVN partition 3.1.6 was used for analysis and NCBI build 36 (hg18) as reference. Patients were screened with brain MRI, and screened for mutation of SOX2, OTX2, RAX and PAX6 using Sanger sequencing. DNA was extracted from lymphoblastoid cell lines from all patients and cultured cells were used for FISH confirmation of CNV. Results A 296.8 Kb deletion in the Yq112ter region was identified and confirmed by FISH on metaphase lymphoblastoid cells using the amplituch subtelomeric Yq/X probe in all 3 boys and healthy father. Three brothers had normal brain MRI as well as negative SOX2, OTX2, RAX and PAX6 mutation analysis by Sanger sequencing. This Yq112ter deletion is within the Pseudo-Autosomal Region PAR2 identical to the Xq28 region. The possible role of the 3 deleted genes from the region in the autistic spectrum disorder will be discussed and the deletion inherited from father cannot clearly exclude its pathogenic role. Discussion: The fact that there is a history of cataract in father’s as well as mother’s family without either parent being affected raises the possibility that the cataract is a separate trait unrelated to the spectrum disorder and inherited from carrier parents in a recessive or even possibly dominant fashion. The deletion CNV identified contains 3 genes and it is difficult without further functional analysis to ascertain causality of this novel CNV in the pathogenesis of the autistic spectrum disorder in the 3 brothers.

2702F
MALFORMATION VARIABILITY ASSOCIATED TO CHROMOSOME TRISOMIES. CLINICAL AND PHENOTYPICAL IMPLICATIONS IN SEVERAL PATIENTS AT A PEDIATRIC HOSPITAL IN MEXICO. M. Barrientos1, J.M. Aparicio-Rodriguez2,3, M.L. Hurtado-Hernandez4, M.A. Cubillo-Leon5, S. Chatelain-Mercado5. 1) Endocrinology; 2) Genetics; 3) Cytogenetics; 4) Rehabilitation Therapy, Hospital para el Niño Poblano, Puebla, Puebla; 5) Biotechnology; 6) Estomatology.

Chromosome trisomies are considered alterations in the chromosome number or structure. A trisomy is therefore a type of polysony in which there are three chromosome copies, instead of the normal two. A trisomy is considered an aneuploidy or abnormal number of chromosomes. There are two different trisomy types: Full trisomy where an entire extra chromosome has been copied. "Partial trisomy" means that there is an extra copy of part of a chromosome. Depending on the chromosome, a trisomy is named as "Autosomal trisomies" (trisomies of the non-sex chromosomes) and "Sex-chromosome trisomies." In this study both Autosomal and Sex-chromosome trisomies are described in different patients, depending on the affected chromosome. Among 4617 chromosomal studies performed during 19 years (from 1992 to 2011), at Hospital Para el Niño Poblano in México, 34.6% (1596 patients) had chromosomal alterations. Among these studies population, a male and female pediatric patients are described, with different chromosome trisomies, were chromosome changes are classified as structural or numeric alterations. All trisomies patients were described in this study analyzing their phenotypical and clinical features, medical treatments and prognosis.

2703W
Congenital primary microcephaly and type B-like brachydactyly, a new syndrome? A. Lavillaureix1, J. Masliah-Planchon1,2, S. Passeraudi1,2, S. Drunat1,2, A. Verloes1,2. 1) Department of genetics, Robert Debré Hospital, Paris, France; 2) INSERM U676, Robert Debre Hospital, PARIS, France.

We report a 10-year-old girl with primary microcephaly and brachydactyly. Microcephaly was suspected during the pregnancy by echography. At birth, at term, head circumference was 29cm (-5 SD), weight 2770g, length 50cm. Pregnancy was normal (no drug nor alcohol) and there were no perinatal problems and no feeding difficulties. She is the first child of unrelated parents none of the members of this family have brachydactyly or microcephaly. She started walking at 8 month-old, first words were around 1 year. She followed mainstream schooling. In 2nd grade, problems with motricity and writing fatigability were observed. Clinical examination at 8 year-old: showed a child of medium stature with generalised amytrophy and microcephaly 41.5cm (-8 SD). Neurological examination was normal. She had facial dysmorphism with upslanted palpebral fissures and microodontia. She has bilateral brachydactyly resembling type B of hands and feet, respecting thumbs but involving halluces, discovered at birth. On X rays, the terminal phalanges were missing on most digits, intermediate phalanges were hypoplastic or aplastic, with anonychia and cone-shaped ephiphyses. MRI revealed a brain of reduced volume but normal structure and gyration. CGH array (180 k), caryotype and mitomicyn-induced chromosome breakage test were normal.

To our knowledge it is the first described case of severe congenital microcephaly with normal IQ associated with brachydactyly and anonychia. This disorder is clinically distinct from Jawad syndrome (due to mutations in RBBP8/CTIP - currently under sequencing), which shows mental retardation and ungual hypoplasia, and Teebi anonychia-microcephaly syndrome, which has much milder anomalies.

2704T
Microrearrangements in individuals within the Holoprosencephaly spectrum. L.A. Ribeiro-Bicudo1, B.F. Gamba1, C. Rosenberg2, A.L.B. da Rocha1, A.L.C. Gaspar1, R.M.C.S. Sandri1, A. Richieri1. 1) Genetics Department, Hospital for Rehabilitation of Craniofacial Anomalies, Bauru, Sao Paulo, Brazil; 2) Human Genome Center, Department of Genetics and Evolutionary Biology, Institute of Biosciences, Universidade de Sao Paulo, Sao Paulo, SP, Brazil.

Holoprosencephaly (HPE) is a malformation sequence where the cerebral hemispheres fail to separate into distinct left and right halves. It can be associated with midline structural anomalies of the central nervous system and/or face. The etiology of HPE is complex, with both environmental and genetic factors being implicated. Numerous different heterozygous mutations have been identified in HPE patients and include missense, nonsense, deletion, and frameshift mutations that are located throughout the gene. Chromosomal abnormalities have been attributed as the main commonly identified cause and high frequency of rearrangements have been reported in studies with array CGH. In the present work we found rearrangements trough Multiplex Ligation-dependent Probe Amplification (MLPA) and arrayCGH analysis in eight non related individuals who presented within the holoprosencephaly (HPE) spectrum previously screened for mutations in some HPE determinant genes such as SHH, GLI2, SIX3, TGIF, and PTCH. Molecular findings showed microdeletions and a microduplication involving SHH, TGIF and ZIC2 genes in five non related individuals, and three individuals presented chromosomal microrearrangements consisting in two duplications, one in a critical region harboring the SIX3 gene (2p21) and other in a region not related with HPE genes (13q14); and a third individual presented a microdeletion (8p23), which is also not associated to HPE phenotype. Because the clinical and genetic heterogeneity existing in patients with HPE and the increasing survival of these patients we concluded, that these analysis are indicated in those cases where the search for mutations in main causative genes were negative providing a better tool for genetic counseling.
2705F
1) Département de Génétique, Robert DEBRE University Hospital and INSERM UMR 676, Paris, France; 2) Département de Gynécologie Pédiatrique, Hôpital Saint-Laurent du Maroni, French Guyana, France; 3) Département de Toxicologie, Rouen University Hospital, Rouen, France; 4) Département de Pathologie, Rouen University Hospital, Rouen, France.

Eating clay ("pemba") during pregnancy is a traditional behavior in the Bushungu population living on the border between the Democratic Republic of the Congo and the French department of Guiana. Clay consist in aluminium silicate. It is a powerful chelator of iron, and this practice (linked to traditional medicine), is responsible for a high incidence of severe anemia of pregnancy in this area. We report on two sibs born to a pemba-eater mother. The first child was born at term with IUGR severe microcephaly. Intracranial "calcifications" were observed by ultrasound screening during the second trimester. CT scan confirmed massive radio-opaque deposits in the brain basis. The clinical diagnosis of TORCH or Acardi-Goutières syndromes were suggested initially. The child survived with major developmental delay. At age 7y, she has an OFC of 39 cm (-10 SD) and a height of 10 cm (-4SD). CSF interferon and TORCH screening were negative. Recurrence of microcephaly during the second pregnancy lead to TOP, after diagnosis of a similar microcephaly.

Neuropathological examination confirmed severe microcephaly, with extensive microcalcifications dispersed throughout the brain. Electron microscopy made it possible to visualize intraneuronal aluminum silicate deposits, resembling aluminnum deposition observed in post-vaccinal myofascitis. The most plausible hypothesis to explain this recurrence is an acquired global IUGR secondary to severe maternal anemia combined with accumulation of exogenous silicates in the neural cells. This appears to be the first description of fetal brain disruption secondary to ingested clay. The syndrome superficially mimics Acardi-Goutières syndrome, and conveys a high risk of recurrence.

2705T

Chromosome aberrations are considered changes in the chromosome number or structure. The etiologic factor is due to gametogenesis inborn error (meiosis) or during the zygote first cellular divisions. It might occur during metaphase from the cellular cycle, where DNA losses are seen (clastogenic processes) due to DNA repair processes deficiency or total absence, among others. Six genetic patients associated to chromosome 4 aberration were analyzed; three Wolf-Hirschhorn syndrome patients, a deletion of long arm 4 chromosome and two 1:4 and 3:4 chromosome translocations among 4617 Karyotype studies performed during 19 years period of time (from 1992 to 2011) at a Pediatric Hospital in Mexico. These chromosome changes are classified as structural alterations where these six patients from different families were chosen to evaluate their clinical characteristics, medical or surgical treatments according to their different genetic aberration.

2706W
Cystic lymphangioma in a 9-year-old boy with Sotos syndrome: a view of the tumoral risk in this overgrowth syndrome. O. Crocco1, G. Fedrasi-zak1, T. Dery2, B. Devauchelle3, V. Strusniki4, L. Burglen4, J.F. Ikoł4, B. Demeer4, M. Mathieu5, A. Leke6, G. Morin1. 1) Genetic department, Amiens University Hospital, Amiens, France; 2) Maxillofacial surgery, Amiens University Hospital, Amiens, France; 3) Otolaryngology, Amiens University Hospital, Amiens, France; 4) Genetic department, Trouseau Hospital, Paris, France; 5) Pathology service, Amiens University Hospital, Amiens, France; 6) Pediatric Reanimation, Amiens University Hospital, Amiens, France.

Background Overgrowth syndromes (OGS) form a heterogeneous group of disorders in which the main characteristic is a weight, height or head circumference over +2 standard deviations. Benign tumours and neoplasms are not exceptional in OGS. Sotos syndrome (SS) (OMIM 117550), one of these OGS, is characterized by cerebral gigantism, distinctive craniofacial appearance, and variable learning disabilities. Haploinsufficiency of the NSD1 gene was identified as the cause of the disorder. From its first description by Sotos in 1964, more than 25 malignancies and 10 benign tumours were reported. The NSD1 gene is localized on chromosome 4q. Molecular and method The case of a patient with SS and cystic lymphangioma is retrospectively described. A PubMed research crossing “Sotos syndrome or cerebral gigantism” and “tumor or tumor or cancer or neoplasm or malignancy” was realized. Results are reported. Case report This male patient has an OFC of 39 cm (-10 SD) and a height of 10 cm (-4SD). CSF interferon and TORCH screening were negative. Recurrence of microcephaly during the second pregnancy lead to TOP, after diagnosis of a similar microcephaly.

Neuropathological examination confirmed severe microcephaly, with extensive microcalcifications dispersed throughout the brain. Electron microscopy made it possible to visualize intraneuronal aluminum silicate deposits, resembling aluminium deposition observed in post-vaccinal myofascitis. The most plausible hypothesis to explain this recurrence is an acquired global IUGR secondary to severe maternal anemia combined with accumulation of exogenous silicates in the neural cells. This appears to be the first description of fetal brain disruption secondary to ingested clay. The syndrome superficially mimics Acardi-Goutières syndrome, and conveys a high risk of recurrence.

2707F
Bench to bedside... The role of clinical genetics in the age of genomic medicine: The shifting paradigm. T. Bardakjian1, A. Šlavoňek2, A. Schneider1. 1) Dept Pedts/Genetics, Albert Einstein Med Ctr, Philadelphia, PA; 2) Dept. Pediatrics, Univ California, San Francisco, CA.

Advances in molecular diagnostics are catapulting genetics into all areas of medicine. The classic paradigm of clinical geneticist as dysmorphologist collecting phenotypic information to identify a syndrome or differential diagnosis and then ordering diagnostic testing is changing. New technology enables identification of gene mutations which are not associated with well-described phenotypes. While new diagnostic technologies offer great potential to provide a genetic diagnosis, when mutations are not accompanied by clinical prognosis or syndromic information the utility of the diagnosis for the patient is in question. The molecular diagnosis is not the end of the involvement of the clinical geneticist but rather a pivotal moment when genetics really becomes the key to best practices of medical care by a knowledgeable team with regular follow-up and monitoring. This paradigm shift is exemplified in the case of gene identification in the rare birth defect anophthalmia/microphthalmia (A/M). Molecular diagnosis is possible in up to 40% of individuals. Numerous eye development genes can now be tested for, but the natural history of these mutations have not been established. SOX2 as an example, was identified as a causative gene in 2005 and was termed SOX2 anophthalmia syndrome. Collection of clinical data has identified a wide variety of ocular findings and systemic issues not initially described. Testing is routinely ordered by a variety of physicians. However, there is limited knowledge of long term issues and it is not possible for most physicians to offer accurate anticipatory guidance for these patients. The A/M Registry at Einstein Medical Center Philadelphia has collected 30 cases of individuals with SOX2 syndrome. Combining these cases with others reported in the literature, the Registry has been able to develop a natural history and phenotype spectrum to provide some guidance. It is important to note that given the infancy of this diagnosis, ongoing follow-up is essential. As more eye development genes are identified and clinical findings broaden the spectrum for each gene, the collection of clinical data over time will be critical to the provision of optimal medical care. Long term clinical and outcome data collection and analysis by clinicians is essential to accurately describe the phenotypes which correspond to identified gene mutations. This will enable more accurate genetic counseling and support for families with rare disorders.
2709W
Targeted next-generation sequencing for the molecular genetic diagnostics of mandibulofacial dysostosis. Y. Kuroda1, I. Ohashi1, T. Saito2, N. Naga1, K. Iida1, T. Naruto1, M. Masuno3, K. Kurosawa1. 1) Division of Medical Genetics, Kanagawa Children’s Medical Center, Yokohama, Kanagawa, Japan; 2) Department of Clinical Laboratory, Kanagawa Children’s Medical Center, Yokohama, Japan; 3) Genetic Counseling Program, Kawan-saki University of Medical Welfare, Kurashiki, Japan.

Mandibulofacial dysostosis (MFD) is a clinically and etiologically heterogeneous group of conditions characterized by significant malar and mandibular hypoplasia. Conducive hearing loss, lower eyelid anomalies, dysplastic ears and cleft palate are frequent associated features. Although many distinct MFD have been described clinically, phenotypic overlap makes it difficult to distinguish syndromic MFD and other craniofacial conditions. Causative mutations have only been identified for syndromic MFD, including TCOF1, POLR1D, and POLR1C in Treacher Collins syndrome, DHODH, SF3B4 in Nager syndrome, and EFTUD2 in FATCO syndrome with microcephaly (MFDM). We performed targeted next-generation sequencing for MFD. Five patient samples were sequenced by MiSeq (illumina). DNA libraries were enriched for sequences by capture-based approach. (HaloPlex, Agilent Technologies) Amplicons were designed for mixed panel of six MFD genes (TCOF1, POLR1D, POLR1C, DHODH, SF3B4, EFTUD2) and 69 ciliopathy genes, covering all coding regions and UTRs, in total, 398,465bp. Mean depth of coverage over all samples was 336x and bases covered by at least 15 reads were 97.5% of CDS. Data were analyzed by BWA ver.6 + GATK pipeline. Calling CNV was based on log ratio and z-score of read depth on each exon. CNVs were found in two of five patients and confirmed by array CGH. No mutation considered likely to be pathogenic was found in all patients. Overall, 2/5 patients (40%) had intragenic deletions as well as pathogenic mutations.

2709F
Worldwide, excluding the nine cases reported in this edition. A novel homozygous mutation [c.G542A] affecting the evolutionary conserved residue p.C181Y was identified at 3q27 in the MASP1 encoding a mannose-associated serine protease 1. The variant was confirmed by Sanger sequencing, segregates with the phenotype in the family and is predicted to be damaging by PolyPhen and SIFT. MASP1 functions as a component of the lectin pathway of complement activation. Mutations in the MASP1 gene and another gene (COLEC11) involved in the same pathway have been associated with human craniofacial malformation indicating an impending role for complement pathway elements in vital developmental processes during embryogenesis. The identified autosomal recessive variant extends further support to this hypothesis.

2712W
Diagnostic Criteria in Gomez-Lopez-Hernandez Syndrome: Contribution of Brazilian Patients. C.H.P. Grangeiro1, L.B. Mesquita1, J.A. Josahkian1, C.M. Leveprost1, M.L.M. Castro1, N.R. Quar-esemin1, L.A.F. Laureano2, A.C. Santos3, J.M. Pinaneto1,3. 1) Serviço de Genética Médica - Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 2) Laboratório de Citogenética - Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 3) Departamento de Genética - Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 4) Divisão de Radiologia do Departamento de Ciência Médica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, Centro de Ciências das Imagens e Física Médica.

Gomez-Lopez-Hernandez syndrome (GLHS) or cerebello-trigeminal-dental dysplasia is a rare neurocutaneous disorder (so far, 34 sporadic cases have been reported), whose etiology is unknown. It is characterized by the triad of thombencephalosynapsis (RES), trigeminal anestesia (TA) and bilateral aplasia of the scapula (SA). Based on the description of these patients and other clinical findings, subsequent reported cases have expanded the spectrum of craniofacial, neurobehavioral and cognitive phenotype. According to data, thombencephalosynapsis and bilateral scapula aplasia constitute obligate criteria, Brachycephaly/turricephaly and midface retrusion are major craniofacial criteria. Strabismus, widely spaced eyes, plagiocepha-ly and lambdoid craniosynostosis are minor craniofacial ones. Motor delay, ataxia, hypotonia, intellectual disability and head shaking or other stereotypic movement are the neurological criteria. After identifying a new patient with GLHS in our service, totaling seven patients (20% of all cases worldwide), we decided to characterize our patients according to the proposed standard criteria and propose some adjustments. Our sample consists of 7 sporadic cases, 4 boys and 3 girls, 2 of them born from consanguineous parents (second cousins). The G-banding karyotypes were normal. We believe that trigeminal anestesia must be considered a mandatory criteria since, not only all patients have this finding, they also had corneal opacity (a complicating factor). All our patients presented broad forehead, upslanting palpebral fissures and low set ears so major craniofacial criteria must include these dysmorphic features. Prognathism and post natal short stature should be considered minor criteria. Neurological findings in our sample are consistent with the literature, including intellectual disability in variable degrees. We did not evaluate the presence of stereotyped head movements. Diagnostic criteria are essential for clinical reasoning construction, search of new patients and discovery of etiologic factors, especially in disorders which underlying mechanisms are not yet known.

2710T

FATCO syndrome is characterized by the presence of Fibulaural Apasia, Tibial Camptomelia (tibial arched), Oligosyndactyly (finger/toes absent and union of one or more) (MIM 246570). Its describe for Hecht and Scott at 1981; but Courtens at 1985 coins the term FATCO syndrome. We present a report of nine children evaluated at the Department of Genetics, Instituto Nacional de Salud del Niño, Lima, Peru; 2) Centro de Investigación de Genética Humana, Lima, Peru. The final diagnosis was MFDM. Patient 2 was older brother of siblings with mandibulofacial dysostosis with microcephaly (MFDM). We performed targeted next-generation sequencing for MFD. Five patient samples were sequenced by MiSeq (illumina). DNA libraries were enriched for sequences by capture-based approach. (HaloPlex, Agilent Technologies) Amplicons were designed for mixed panel of six MFD genes (TCOF1, POLR1D, POLR1C, DHODH, SF3B4, EFTUD2) and 69 ciliopathy genes, covering all coding regions and UTRs, in total, 398,465bp. Mean depth of coverage over all samples was 336x and bases covered by at least 15 reads were 97.5% of CDS. Data were analyzed by BWA ver.6 + GATK pipeline. Calling CNV was based on log ratio and z-score of read depth on each exon. CNVs were found in two of five patients and confirmed by array CGH. No mutation considered likely to be pathogenic was found in all patients. Overall, 2/5 patients (40%) had intragenic deletions as well as pathogenic mutations.

FATCO syndrome is a rare neurocutaneous disorder (so far, 34 sporadic cases have been reported), whose etiology is unknown. It is characterized by the triad of thombencephalosynapsis (RES), trigeminal anestesia (TA) and bilateral scapula aplasia (SA). The apparent clinical characteristics overlap, but do not identify solely, with the individual Malpuech, Michaels, Mingarelli or Carnevale syndromes, or what has been collectively referred to as the 3MC syndrome; hence, the reference to the phenotype as Multiple Congential Anomaly syndrome. The family was studied by homozogosity mapping, and Whole Exome Sequencing of a single affected individual performed on ABI SOLID4. A novel homozygous mutation [c.G542A] affecting the evolutionary conserved residue p.C181Y was identified at 3q27 in the MASP1 encoding a mannose-associated serine protease 1. The variant was confirmed by Sanger sequencing, segregates with the phenotype in the family and is predicted to be damaging by PolyPhen and SIFT. MASP1 functions as a component of the lectin pathway of complement activation. Mutations in the MASP1 gene and another gene (COLEC11) involved in the same pathway have been associated with human craniofacial malformation indicating an impending role for complement pathway elements in vital developmental processes during embryogenesis. The identified autosomal recessive variant extends further support to this hypothesis.
2713T

We present a newborn girl with brain malformations, bilateral Peters anomaly and multiple intestinal atresias and we compare this case with the only other similarly affected case reported in the medical literature (Shanske AL et al, 2002). Our case was born at 31.3 weeks vaginally without complications to a 34yo primigravida mother who was prenatally followed for mild fetal ventriculomegaly. The family history was not significant. Our clinical examination revealed a very small head circumference and a birth weight in the 50th centile for gestational age. She had short palpebral fissures, two skin tags in place of the left tragus, a skin tag on the left cheek, corneal opacities and normal tone and activity. Feeding intolerance prompted an abdominal x-ray that suggested proximal jejunal atresia. Exploratory laparotomy revealed proximal jejunal atresia 5 cm distal to the ligament of Treitz, multiple discrete atretic segments of jejunum, a 5 cm segment of atretic distal ileum, and colonic atresia in the mid transverse colon. The atretic segments were resected, a mucous fistula constructed, and a gastrostomy tube placed. Brain MRI demonstrated bilateral schizencephaly communicating with large biperiartal extra-axial cysts, partial absence of corpus callosum, ventriculomegaly, small areas of polymicrogyria and generally immature sulcation and myelinlation pattern for age. Ophthalmology concluded that the corneal opacities were part of Peters anomaly. Whole genome chromosome SNP microarray using the Affymetrix Cytoscan HD platform was reported normal. DNA sequence test of the exons, flanking regions and exon-intron boundaries of the B3GALT1 gene to evaluate for Peters Plus syndrome revealed no mutation. The atresias started on day 12, and episodes of respiratory depression and bradycardia on day 19. She expired on day 20. Shanske AL et al, 2002 described a male newborn with extensive neuronal migration defect, bilateral Peters anomaly, multiple stenotic and atretic lesions in the jejunum and normal karyotype. At 21 months, he was severely microcephalic and developmentally delayed. The authors concluded that the abnormalities were due to a vascular disruption sequence. Both our case and the case reported in the medical literature present with migration defects, bilateral Peters anomaly and multiple intestinal atresias. This phenotype does not fit a recognizable pattern of malformation and we suggest that it may represent a new syndrome.

2714F
Richieri-Costa and Pereira Syndrome: severe phenotype. S. RASKIN2, M. SOUZA1, M.C. MEDEIROS3, M. MANFRON2, D.C. CHONG E SILVA2. 1) Group for Advanced Molecular Investigation, Graduate Program in Health Science, Health and Biosciences School, Pontificia Universidade Católica do Paraná (PUCPR), Curitiba, Parana, Brazil; 2) Hospital Pequeno Príncipe, Curitiba, Parana, Brazil; 3) Health and Biosciences School, Pontificia Universidade Católica do Paraná (PUCPR), Curitiba, Parana, Brazil.

Richieri-Costa/Pereira syndrome (RCP, OMIM 268305) is a rare autosomal recessive disorder characterized by short stature, Robin sequence (micrognathia, glossoptosis, and cleft palate), cleft mandible and limb malformations. We report on a new case with a very severe phenotype. The patient was the first born of a 1st cousin consanguineous marriage, with no familial history of skin disorders. Prenatal care and ultrasound exams showed no fetal morphological alterations. Delivery was at 38.5 weeks of gestation, the infant weighed 2,805 g, was 44 cm long, with a OFC of 31 cm and an APgar score of 5 at 1 and 5 minutes, respectively. Malformations included: radial dysmelia, and finger anomalies, club feet, short limbs, toe anomalies, micrognathia, thoracic deformity with severe sterno-clavicular chondral bilateral dysplasia, fatty hyperplasia of the anterior and posterior surface of the neck, webbed neck, low-set ears, ear deformities, and cranio-facial dysostosis evolving to craniosynostosis. He had prolapse of base of tongue over the larynx and agenesis of the epiglottis; facial CT scan showed mandibular agenesis and deviated nasal septum. Limb radiographs showed radial agenesis, hypoplastic ulna, malformations of hands and fingers, hyperplasia of the tibia, agenesis of fibulae and toe deformities. Tube feeding was necessary on the time of birth, due to anomalies of the air ways. Three-dimensional CAT scan showed agenesis of the anterior mandible arch and epiglottis. Cranial CAT scan showed prominence of extra-axial space of the frontal convexity. Alterations were observed in skull formation with overgrowth of the lambdoid sutures. To the present, 32 Brazilian and one non-Brazilian case have been described, with a great variability in the expresion, but none reported the same degree of severity as the present case. Absence of Robin sequence and cleft mandible have been previously reported, but the present case, particularly in a 1st cousin, a severe phenotype is unexpected. The recent finding of a causative genetic alteration at the 5 UTR of the EIF4AE gene as the mutation leading to RCPs may generate information on the variability of expression and further elucidate the basis of severe phenotypes such as that reported in the present study.

2715W
Minor facial malformations in relatives of patients with Goldenhar syndrome. P. Santos1, S. Oliveira2, H. Saffatle3, M. Cordoba1, F. Ferrari2, J. Mazzeu4. 1) Programa de Pos-graduação em Ciencias da Saude, Universidade de Brasilia, Brasilia, DF, Brazil; 2) Laboratorio de Genetica, Departamento de Genetica e Morfologia, Instituto de Ciencias Biologicas. Universidade de Brasilia, Brasilia, Brazil; 3) Ambulatorio de Genetica, Hospital Universitario de Brasilia, Brasilia, DF, Brazil; 4) Programa de Pos-Graduacao em Ciencias Genomicas e Biotecnologia, Universidade Catolica de Brasilia, Brasilia, DF, Brazil.

Goldenhar syndrome (GS), also known as oculo-auro-vertebral syndrome, is a congenital defect from abnormal development of first and second branchial arches. Its etiology includes unknown genetic factors, different chromosome aberrations and environmental factors, like maternal vasoactive medication, thalidomide and maternal diabetes. Several candidate genes have been proposed but none have been confirmed as causative of the phenotype. It’s a disease with genetic heterogeneity and variable expressivity, commonly underdiagnosed. The majority of cases are sporadic so an environmental contribution to the phenotype cannot be excluded. Only a few familial cases have been reported so far. We studied a large family with GS, with three patients presenting classic signs of Goldenhar syndrome such as facial asymmetry, hemifacial microsomia, microtia or anotia and pre-auricular tags. Other family members exhibited mild malformations: mild facial asymmetry, hyperextended ear helix and pre-auricular pits. These signs though common in the normal population when observed in family members of Goldenhar syndrome patients may be indicative of an inherited form of the syndrome. Financial support: CAPES, FAPDF/PPUS.

2716T
Wiedemann-Beckwith syndrome associated with pre- and postnatal supraventricular tachycardia. M. Willems1, F. Brioude2, S. Guillaumont3, P. Amelro2, M. Vincenti2, O. Pidoux1, N. Fries4, L. Begue5, C. Dumont6, P. Sarad1, P. Blanchet1, L. Pinson7, E. Haquett1, J. Puechbery8, G. Lefort7, C. Coubes1, I. Netchine1, D. Genevieve1. 1) Genetics Dept, INSERM U844, Hopital Arnaud de Villeneuve, Montpellier, cedex 5, France; 2) Explorations Fonctionnienelles Endocrinienes, Hospital Armand Trousseau, Paris, cedex 12, France; 3) Service de Cardiopadiatre, hopital Arnaud de Villeneuve, Montpellier, cedex 5, France; 4) Service de Gynecologie Obstetrique, hopital Arnaud de Villeneuve, Montpellier, France; 5) Service de Pédiatrie Nextonatologie, hopital Arnaud de Villeneuve, Montpellier, cedex 5, France.

We describe a baby who has Wiedemann Beckwith syndrome due to ICR2 (CDKN1C/KCNQ1OT) loss of methylation, associated with severe fetal and postnatal supraventricular tachycardia. The patient is the third child born from unrelated parents after in vitro fertilization with donor sperm. It was a bichorial biamniotic pregnancy with an early vanishing twin. Prenatal ultrasound screening at 14 WG revealed an omphalocele, leading to perform an amniocentesis. Karyotype analysis was normal. At 29WG, the mother was referred to the Genetic Department for the association of fetal macrosomia, macroadenotrophy, hepatomegaly and hydramnios. Wiedemann-Beckwith syndrome was suspected. At 30 WG, fetal supraventricular tachycardia was identified, requiring a maternal treatment with digoxin and amiodaron which is ongoing at four months of age. Ectopic atrial tachycardia was reported only in two patients of Goldenhar syndrome patients may be indicative of an inherited form of the syndrome. Financial support: CAPES, FAPDF/PPUS.
2717F
Amyoplasia with congenital eye malformations and wrinkled skin: a new syndrome. D.F.G.J. Wolthuis1, E.V. van Asbeck2, H.C. Andersson3, E. Morava-Kozicz, MD, PhD,1,1) Pediatrics, Radboud University Nijmegen Medical Centre, Nijmegen, Gelderland, Netherlands; 2) Human Genetics, Tulane School of Medicine, New Orleans, LA.
Amyoplasia is characterized by the congenital absence of muscle tissue, and the most common underlying condition in arthrogryposis multiplex congenita. Patients present with typical symmetrical contractures, including internal rotation of the shoulder, extension of the elbow, flexion of wrists and hands, equinovarus feet and variable contractures of knees and hips with additional supra-articular dimples. Midline hemangiomas are also a common. All published amyoplasia cases are sporadic. Compared to the genetically heterogeneous distal arthrogryposis (DA) syndromes, no associated organ malformations have been described in amyoplasia patients. DA syndromes describe conditions of contractures in 2 or more body parts in the absence of a primary neurologic or muscular disease. We evaluated 2 patients diagnosed with muscle biopsy and imaging-proven amyoplasia in combination with unusual symptoms. Both patients showed eye abnormalities in addition to absent muscle: patient 1 had macular abnormalities and patient 2 had cataract. Feeding problems were present in both patients as well as spontaneous bone fractures, excessive skin in the nuchal region and typical facial features. Patient 1 also had diaphragmatic paresis. Amyoplasia has never been described in combination with eye abnormalities in the medical literature. However, retinal abnormalities are a feature that is commonly seen in DA, especially in type 5 and 5D. In DA5D a genetic defect was identified in ECEL1, encoding endothelin-converting enzyme-like 1. Mutation in this endopeptidase leads to reduced terminal branching of motor neurons resulting in decreased development of neuromuscular junctions, which was ruled out in our patients. The underlying pathomechanism in amyoplasia has not yet been discovered. Infectious and autoimmune causes, muscle dystrophy, spinal motor neuropathy and congenital disorders of glycosylation, affecting both muscle and eye development, were ruled out in our patients. Immune histology of extracellular matrix components, however, was abnormal. We suspect the role of genetic factors, like genes involved in collagen synthesis, affecting the extracellular matrix, and development of both the muscle tissue and eye structures, underlying the unique phenotype in our patients.

2719T
A case with single deletion of 17q21.31 involving KANSL1 gene and phenotype of CHARGE association. ym. chan1, kw. choy2, ty. leung1, ca. bacino1. 1) Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong , Hong Kong; 2) Department of Molecular and Human Genetics Baylor College of Medicine.
We describe a single case with prenatal ultrasound showing increased fetal nuchal translucency and features suggestive for CHARGE association with a small deletion in 17q21.31 involving the KANSL1 gene. On routine first trimester screening ultrasonography performed at 13 weeks of gestation, nuchal translucency was increased to 4.53mm. Chorionic villi sampling was performed. Conventional karyotype was normal. Array CGH detected 573kb single deletion at 17q21.31 involving the KANSL1 gene region, but not included other genes like MAPT and CRHR1. The clinical significance of the deletion at this region was not well known at that juncture. Anomaly scan was unremarkable. Multiple abnormalities were noted after delivery including bilateral membranous choanal atresia, bilateral coloboma involving fovea and optic nerve, bilateral malformed pinna with bilateral severe-profound hearing loss, patent ductus arteriosis compatible with CHARGE association. CHD7 sequencing analysis was normal with no CHD7 mutation identified. Patients with KANSL1 related intellectual disability may present with dysmorphic features resemble those seen in CHARGE association. However, no reports of choanal atresia or colobomas were previously reported so this is the first report for these malformations in a patient with a KANSL1-related intellectual disability deletion.

2720F
Femoral-facial syndrome: long term follow-up and associated array CGH abnormalities. A. Jacquinet1, H. Valdés-Socín1, C. Liblouille1, J.H. Caberg1, A. Verloes1,2, 1) Pediatrics, Radboud University Nijmegen Medical Centre, Nijmegen, Gelderland, Netherlands; 2) Clinical Genetics, Erasmus MC-Sophia, Rotterdam, Zuid-Holland, Netherlands.
The femoral-facial syndrome is usually sporadic and its aetiology remains unknown. Non-genetic factors as maternal diabetes mellitus have been associated. Reports of familial cases have otherwise suggested autosomal dominant inheritance. We report the 20 years clinical follow-up of a girl with femoral-facial syndrome diagnosed at birth. Recently, array CGH investigation identified a 1400 kb duplication at 9q31.1, including the gene SMCM2, and a 343 kb deletion at 12q24.33 including the genes CHFR, ZNF26, ZNF140, ZNF10 and ZNF268. Moreover, the patient presents a Mayer-Rokitansky-Kuster-Hauser syndrome diagnosed at puberty. Femoral-facial syndrome and Mullerian agenesis may reflect different defects in the primary axial mesodermal development, being the consequences of same environmental or/and genetic factors during blastogenesis. Among these genetic factors, we suggest the possible involvement of the two copy number variants reported here.
Identification of Mosaic Activating Mutations in Overgrowth Syndromes Using a Customized Next Generation Sequencing Panel on both Prenatal and Postnatal Samples. L. Liu1, F. Chang1, E. Fang1, G. Zhang1, M.M. Li1,2. 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dan Duncan Cancer Center, Baylor College of Medicine, Houston, TX.

Overgrowth syndromes are genetically heterogeneous diseases caused by both germline and somatic mutations of different genes. Recent studies have shown that a group of overgrowth syndromes, such as CLOVES and Proteus syndromes, are caused by postzygotic activating mutations in the genes involved in the PI3K-AKT signaling pathway, such as PIK3CA and AKT1. In addition, both germline and somatic mutations of PIK3CA, AKT1, GNAS, or GNAS-related genes have been reported in two overlapping disorders, megalencephaly-polydactyly-hydranencephaly (MPH) and megalencephaly-cavitary malformation (MCP). Non-syndromic overgrowth features, such as isolated macrodactyly, have also been reported to be associated with PIK3CA mutations.

Due to the low-abundance nature of these mutations, routine Sanger sequencing often yields negative results. We developed a next generation sequencing (NGS) test that targets all known mutations in multiple genes involved in the PI3K-AKT pathway. For differential diagnosis purposes, we also included a few other genes associated with overgrowth syndromes, such as PTEN and GNAS.

Four patients including two prenatals, an amniotic fluid and a POC samples, and two postnatal cases suspected of CLOVES or Proteus syndromes were tested in our laboratory using the NGS panel. All four cases are positive for PIK3CA somatic mutations including one G542K, one H1047R, and two H1047L mutations. These mutations are only present in the affected tissues. In the amniotic fluid case, the G542K mutation was present in the DNA extracted from cultured amniocytes but negative in uncultured fluid of the same specimen. These results suggest that the mutant allele is present in the direct amniotic fluid sample at a very low frequency beyond the detection limit of the test. The activating mutation may render growth advantages to the cells carrying the mutation in culture, resulting in the enrichment of mutant allele. To the best of our knowledge, this is the first case of prenatatal diagnosis of CLOVES syndrome.

Our experience demonstrates that cultured amniocytes can be used for the detection of fetal mutations associated with syndromes like CLOVES. The custom-designed NGS panel shows high accuracy and sensitivity for the detection of causal mutations in the overgrowth syndromes and facilitates clinical diagnosis both prenatally and postnatally.

Association of mosaic activating mutations with PIK3CA, AKT1, GNAS-related genes, and other genes associated with overgrowth features was performed in our laboratory using the NGS panel. All four cases are positive for PIK3CA somatic mutations including one G542K, one H1047R, and two H1047L mutations. These mutations are only present in the affected tissues. In the amniotic fluid case, the G542K mutation was present in the DNA extracted from cultured amniocytes but negative in uncultured fluid of the same specimen. These results suggest that the mutant allele is present in the direct amniotic fluid sample at a very low frequency beyond the detection limit of the test. The activating mutation may render growth advantages to the cells carrying the mutation in culture, resulting in the enrichment of mutant allele. To the best of our knowledge, this is the first case of prenatatal diagnosis of CLOVES syndrome.

Our experience demonstrates that cultured amniocytes can be used for the detection of fetal mutations associated with syndromes like CLOVES. The custom-designed NGS panel shows high accuracy and sensitivity for the detection of causal mutations in the overgrowth syndromes and facilitates clinical diagnosis both prenatally and postnatally.

Novel GATA4 promoter polymorphism associated with congenital heart disease in south Indian patients. S. Mattapally1, KS. Murthy2, S. Nizamuddin3, K. Thangaraj4, SK. Banerjee5. 1) Division of Pharmacology, Indian Institute of Chemical Technology (IICT), Hyderabad,500 007, India; 2) Innova Children's Heart Hospital, Tarnaka, Hyderabad, India; 3) Centre for Cellular and Molecular Biology, Habsiguda, Uppal Rd, Hyderabad 500 007, India.

Background: Congenital heart diseases (CHDs) usually refer to abnormalities in the heart's structure or function that arise before birth. Although the exact mechanism behind this cardiac abnormality is not known, transcription factors play an important role in embryonic heart development. GATA4 is one of the candidate transcription factors and GATA4 mutation may lead to different types of CHD such as ASD, VSD, TOF and Single ventricle (SV). The aim of this study is to find the genetic association of CHD with GATA4 mutations from south Indian CHD patients. Method: GATA4 gene was genotyped in 100 CHD patients (ASD, VSD, TOF and SV) and 200 control samples in a case control study using sanger's di-deoxy chain termination method. Results: In this study we identified one 5’ UTR (promoter region -490 to 100 bp) mutation i.e., 620 C>T (rs16777615, p>0.0007) and two coding region mutations i.e., c.1734 C>A (Pro394Thr) and c.1827 G>A (Asp425Glu). We also found five 3’UTR mutations i.e., 2400 T>C (rs844862), 2415 T>C (rs904018), 2470 A>C (rs840291), 2446 C>G (rs12825), 3139 A>T (rs12458). All mutation present in the CHD patients are absent in 200 healthy volunteers. Our ‘in silico’ data also provide evidence that all mutations reported above are pathological or alter the gene expression through micro RNA binding. Conclusion: The present study found that GATA4 genetic variations are associated with CHD in South Indian patients. Our bioinformatics study provides further evidence that those GATA4 mutations observed in Indian patients mutations may alter the function of the transcription factor binding and micro RNA binding, may lead to disease.
2725T A new case of Crane-Heise syndrome with comparative review of literature. A. Hantel1, K. Fay2, M. Costadia2, R. Lebel1. 1) SUNY Upstate Medical University, Syracuse, NY; 2) Pathology Associates of Syracuse, Crouse Hospital, Syracuse, NY.

Crane-Heise syndrome is a lethal constellation of dysmorphic features described in only 9 cases (first reported in 1981). The syndrome is notable for 1) poorly mineralized calvarium, cleft lip, cleft palate, low-set ears, hypertelorism, and 3) skeletal abnormalities including dysplastic clavicles, vertebral anomalies, and talipes equinovarus. We present the tenth case of this syndrome. The G4P2012-2022 Caucasian 35-year-old woman had one ectopic pregnancy and two surviving daughters; one said to be affected by autism and seizures, the other by ADHD and Asperger syndrome. Ultrasound at 7 weeks provided gestational dating. Repeat ultrasound at 13 weeks revealed gastrochisis, scoliosis, dangling choroid plexus, and left clubbed foot. Ultrasound at 16 weeks confirmed those observations and added cleft lip and palate, and bilateral ventriculomegaly. Amniocentesis revealed a 46,XX karyotype and normal oligo-microarray. Alpha-fetoprotein was elevated with a faintly positive acetylcholinesterase band. Ultrasound at 21 weeks showed fetal demise, apparently three weeks earlier by measurements, and delivery was induced. Autopsy revealed many features seen in Crane-Heise syndrome: markedly undermineralized calvarium, cleft lip, cleft palate, hypertelorism, soft tissue syndactyly of one hand, low-set ears, and left clubbed foot. Our patient also had a large gastrochisis with extracorporeal viscera; abdominal wall defects have been described in 2 of the 9 previous cases. The patient had renal hypoplasia, not described in previous cases. There was also scoliosis, a narrow thorax, and markedly hypoplastic lungs (each described in only one previous case). Features commonly reported in the previous cases, but absent in the present case are: IUGR, agenesis of vertebral bodies, micrognathia, absent corpus callosum, absent pubic bone, cardiac abnormalities, and hypoplasia of phalanges. This fetus had findings that did not fit together unless one considered a very rare syndrome. Leprosy Monomyelopathy Database. Amniocentesis revealed many features seen in Cran-Heise syndrome as a strong candidate. We propose that Crane-Heise syndrome should be considered in the differential diagnosis for prenatally discovered gastrochisis. There is strong evidence that this syndrome has a familial recessive etiology, and our case does not contradict this hypothesis.

2726F Pathogenic CNVs and causative gene analysis by SNP arrays as the third screening for 846 patients with intellectual disability and multiple congenital anomalies of unknown etiology. D.T. Uehara, S. Hayashi, J. Inazawa. Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan.

In order to identify genetic factors responsible for congenital disorders of unknown etiology in the Japanese population, we have investigated 846 subjects presenting with clinically uncharacterized multiple congenital anomalies and intellectual disability for eight years. We first performed a two-stage screening using two types of in-house bacterial artificial chromosome (BAC)-based arrays, which have allowed the identification of pathogenic copy number variants (CNVs) in 130 cases (20%). Next, we performed a third screening in 432 negative cases using a single nucleotide polymorphism (SNP) arrays platform (Illumina HumanOmniExpress BeadChips) to identify smaller causative CNVs undetected in the previous screenings. Besides the high resolution, this type of array has an additional advantage of detecting uniparental disomy (UPD) through the analysis of copy-neutral loss of heterozygosity (LOH). A preliminary analysis of the results revealed around 10% of the cases with CNVs comprising known pathogenic genes, for example, a deletion in the NRXN1 gene. So far, we have identified five cases with possible pathogenic CNVs that might reveal novel candidate genes for intellectual disability, and one case with uniparental disomy affecting chromosome 20. We are currently focusing on these cases by carrying out functional analyses, and we also plan to further elucidate the mechanisms that originated these rearrangements. Our goal is to finally determine the causative genes delineating these pathways through a reliable correlation between genotype and phenotype.

2727W The oculoauriculovertebral spectrum: refining the estimate of birth prevalence. M.T. Gabbett. Genetic Health Queensland, Royal Brisbane & Women’s Hospital, Herston, Queensland, Australia.

The oculoauriculovertebral spectrum (OAVS) is a pattern of congenital malformations characterized by hemifacial microsomia and/or auricular dysplasia. However, the birth prevalence of OAVS is poorly characterized. Figures ranging from 1 in 150,000 through to 1 in 5,600 can be found in the literature. This study aims to evaluate the reasons behind such discrepant figures and to refine the estimated birth prevalence of OAVS. Published reports on the incidence and prevalence of OAVS were systematically sought after. This evidence was critically reviewed. Data from appropriate studies was amalgamated to refine the estimate of the birth prevalence for OAVS. Two main reasons were identified why birth prevalence figures for OAVS are so highly discrepant: differing methods of case ascertainment and the lack of a formal definition for OAVS. This study refines the estimate of birth prevalence for OAVS to between 1 in 40,000 and 1 in 30,000. This number needs to be confirmed in a prospective study using a formally agreed-upon definition for OAVS.

2728T Initial data for benign CNVs distribution in Bulgarian patients. S.P. Hadzhidekova1, D.M. Avdjieva-Tzavella2, B.B. Rukova1, D.V. Nesheva1, R.S. Tinceva2, D.I. Toncheva1. 1) Medical Genetics, Medical University - Sofia, Sofia, Sofia, Bulgaria; 2) Section of Clinical Genetics, State University Pediatrics Hospital ‘Queen Evdokia’, Medical Faculty, Medical University-Sofia, bul. ‘Iv. Geshov’ 11, Sofia 1660, Bulgaria.

Introduction: Molecular karyotyping is coming up as an extremely suitable method for genetic diagnosis of patients with unclear dysmorphic syndromes and intellectual disability. In this study we present our results from microarray analysis of 52 patients with developmental delay and congenital malformations. Methods: Oligo array-CGH (BlueGnome CytoChip oligo 2x 105K, v1.1, 35kb backbone resolution) was applied in 52 patients with developmental delay and multiple congenital anomalies. Results: A total of 247 CNVs were detected, of which 15 pathogenic, 108 with unknown clinical significance and 124 benign - mean number of CNVs per patient - 4.5. All pathological findings were validated by FISH analysis. In addition, the majority of the patients tested (41 patients) showed normal variations in the number of copies and variations of unknown clinical significance (34 patients). Analyses of the type and distribution of the different variations was performed and the clinical significance of variants of unknown nature was discussed. Conclusion: Our results show the advantages of high resolution microarrays for clinical diagnosis of patients with intellectual disability and congenital malformations, but also highlight the need for extensive population studies revealing the molecular nature and clinical significance of different CNVs and the necessitate for creation of detailed maps of variations in the Bulgarian population. This would facilitate the interpretation of unknown genomic imbalances in clinical aspect. Besides, it would help the widespread introduction of CGH microarray in diagnostic practices - postnatal and prenatal genetic diagnosis. Acknowledgements: Grant 0276-21.12.2008, National Science Fund, Bulgaria.
2729F
Novel autosomal-recessive syndrome with short stature, distinct facial appearance, myopia, retinitis pigmentosa, bilateral hearing loss, and mild intellectual disability. E. Schrock1, T.M. Neuhammer2, I. Neuhammer1, A. Bierer3, B. Novotna4, N. Di Donato1, 1) Institut fuer Klinische Genetik, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Germany; 2) Medizinisch Genetisches Zentrum, Munich, Germany; 3) MVZ Prof. Neuhammer, Munich, Germany; 4) Gemeinschaftspraxis fuer Humangenetik, Dresden, Germany; 5) Sozialpädiatrisches Zentrum am Universitätsklinikum Carl Gustav Carus, Dresden, Germany.

Retinal anomalies in combination with progressive hearing loss accompany a variety of genetic syndromes. We report here on three patients with retinitis pigmentosa, bilateral hearing loss, peculiar facial phenotype and an unusual combination of other clinical findings incompatible with any of the known genetic conditions. Patient 1 (3 years) presented with short stature (-3, 5 SDS), progressive bilateral hearing loss, myopia (-3dp) and progressive bilateral hearing loss. Her motor development is mildly delayed. Patient 2 (40 years) is the paternal aunt of patient 1. Her brother, the father of patient 1, is unaffected. Like her niece patient 2 shows a short stature (-3, 5 SDS), brachydactyly and strikingly similar facial anomalies. She has progressive bilateral hearing loss with deafness at the age of 35 years. The complex eye involvement includes progressive myopia, retinitis pigmentosa (onset in the 2nd decade), glaucoma and corneal dystrophy (both started in the 4th decade). Patient 2 attended a regular school but needs support in her daily life. Both patients have normal findings in their brain MRI. Patient 3 (29 years) is a male patient, not related to patients 1 and 2. His family history is unremarkable. He presented with short stature, brachydactyly, progressive bilateral hearing loss, high myopia and retinitis pigmentosa (onset in the 2nd decade). His facial minor anomalies are remarkably identical to patients 1 and 2. The only additional clinical finding not present in the first family is alopecia areata of the scalp. Patient 3 shows a severe intellectual disability. Comprehensive genetic tests (conventional and molecular karyotyping in all patients; sequencing of LTBP2, ADAMSTS17) were negative. Taken together we report on an apparently new genetic syndrome.

2730W
The use of exome sequencing to disentangle complex phenotypes. H.J. Williams1, C. Bacchelli1, J. Hurst2, F. Lescai3, L. Ocaka1, C. James1, C. Pao1, E. Rosser4, P. Beales1, 1) The Centre for Translational Genomics - GOSGene, Institute of Child Health, UCL, London, UK; 2) Royal Free Campus, UCL Medical School. London, UK; 3) Respiratory and General Paediatrics, Bart's and The London Children's Hospital, London, UK; 4) Clinical Genetic, Great Ormond Street Hospital, London, UK.

The success of whole-exome sequencing (WES) to identify mutations causing single gene disorders has been well documented. However, WES has had limited success in the identification of more complex phenotypes resulting from the disruption of multiple genes. We describe a family where two offspring from healthy consanguineous parents present a complex congenital nonsyndromic phenotype consisting of peripheral neuropathy and bronchietasis that has not been described previously. Through the use of WES we were able to simplify this complex phenotype and identify a causative mutation (R1070X) in the gene PRX, a gene previously shown to cause Charcot-Marie-Tooth Syndrome 4F and Dejerine-Sottas syndrome when this mutation is present in a homozygous state. For the bronchietasis phenotype there was no single mutation or compound heterozygosity identified which was deemed to be causal, reflecting the heterogeneous nature of this phenotype. This study highlights the potential utility of WES to disentangle complex phenotypes where multiple contributing loci in combination with environmental factors make it difficult to assign a clinical diagnosis; this then has further implications regarding the clinical management and use of therapeutics for such patients. In conclusion we show that WES has the power to improve patient diagnosis and therapy by disentangling complex phenotypes through the identification of causative genetic mutations for distinct clinical disorders that were previously masked.

2731T
Association study of genetic polymorphisms in DNA repair genes APE1/Ref-1 and DNA oxidative damage with the risk of neural tube defects. J. Wang1, X. Han2, J. Guo2, X. Wang2, F. Wang2, C. Ji2, Z. Guan1, O. Xie1, Z. Zhu1, B. Niu1, T. Zhang2, 1) Department of biotechnology, Capital Institute of Pediatrics, Beijing, China; 2) Department of molecular immunology, Capital Institute of Pediatrics, Beijing, China.

Neural tube defects (NTDs) are one of the most common human birth defects. Folate deficiency is closely related to NTDs. However, the mechanism remains unclear. Folate can prevent the effect of homocysteine on oxidative stress. Apurinic/apyrimidinendonuclease 1/redox-factor 1 (APE1/Ref-1) plays critical roles in DNA oxidative damage repair, oxidative stress. Previous study found the gene polymorphisms were associated with NTDs in a Californian population. In order to investigate the association of APE1/Ref-1 gene polymorphisms and oxidative stress with NTDs among the folate deficiency population in the high-risk area, a case-control study of 335 NTDs fetuses and 336 normal fetuses was conducted in Lviang areas of Shanxi province with a high prevalence of NTDs. Total 36 single nucleotide polymorphisms (SNPs) in APE1/Ref-1 gene were genotyped by Sequenom MassARRAY Genotyping. The Superoxide dismutase (SOD) activities of fetal liver and maternal plasma were detected by the Cu-Zn/ Mn-SOD assay kit and the numbers of apurinic/apyrimidinic sites (AP sites) in DNA of fetal brain tissues were measured by the DNA Damage Quantification Kit. The allele and genotype frequencies of 36 polymorphisms showed no statistically associated with NTDs. After stratifying subjects by NTD phenotype, we observed 4 polymorphisms statistically associated with NTDs subtypes. As to rs1130409, compared with TT genotype, GG genotype of was associated with the decreased risk for spina bifida, in female TG genotype of was associated with the decreased risk for cephalic malformations compared with those harboring the TG genotype. Allele C of rs3136817 was associated with an increased risk for single encephalocele compared with those harboring allele T. In male, allele T of rs77794916 had a 1.547-fold increased risk for NTDs compared with those harboring allele C. In female spina bifida. The results of this study imply that the APE1/Ref-1 gene polymorphisms were correlated with susceptibility to NTDs, oxidative stress and DNA oxidative damage was associated with NTDs in a high-risk area of China.
2732F
Whole exome sequencing of a girl with Rubinstein-Taybi syndrome. H. Yoo1,2, K. Kim1, I. Kim1, S. Rha1, J. Park1, S. Kim2, N. Kim2. 1) Department of Psychiatry, Seoul National University Hospital, Seongnam, Gyeonggi, Korea; 2) Seoul National University College of Medicine, Seoul, Korea; 3) Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea; 4) Softmatter Research Center, Goyang, Gyeonggi, Korea; 5) Department of Pharmacology, Eulji University College of Medicine, Daejeon, Korea.

Objectives: The Rubinstein-Taybi syndrome (RSTS) is a rare condition with a prevalence of 1 in 125,000–300,000, with dysmorphic features of face, hands & feet (Rubinstein et al., 1963). The genetic mutations of RSTS are not confirmed, though CREBBP or EP300 mutations have been reported (Tsai et al., 2011). The purpose of this study is to evaluate the genetic causes of RSTS with Whole Exome Sequencing (WES). Methods: A 6-year-old Korean girl with RSTS was clinically phenotyped and behaviorally assessed with WPPSI, VABS, Leiter-R,ADOS, ADI-R, CBCL & Pittsburgh Sleep Quality Index (Park et al., 1996; Sparrow et al., 1994; Gale et al., 1997; Lord et al., 1994; LeCouteur et al., 2003; Achenbach et al., 1991; Buysse et al., 1989). Blood samples are drawn from the proband and both biological parents. Read mapping, duplicate removal, local re-alignment, SNP and short indel genotyping have been performed by BWA, Picard, and GATK. Gene annotation and variant filtering is done in in-house bioinformatics pipeline and control database comprised of 54 normal Korean individuals. Results: 1) Phenotype: The proband is the only child of a healthy non-consanguineous couple. Her features include: 1) Typical facial dysmorphism: highly arched eyebrows, long eyelashes, down-slanting palpebral fissures, beaked nose with the nasal septum extending below the alae, dental crowding & micrognathia; 2) Broad & angularized thorax & halluces, thickening of the soft tissue of the phalanges & persistent fetal pads; 3) Microcephaly, growth retardation (head circumference, Wt & Ht<3 percentile); 4) Congenital cataract; 5) Skin problems; 6) Swallowing difficulty; 7) EEG abnormality; 8) Mental retardation (IQ=37, PIQ=45); 9) Autism spectrum disorder. We observed de novo mutation of CREBBP (c.1415C>T). (c.323G>A) & (c.323A>G): We observed de novo mutation of CREBBP, EP300 (c.1415C>T). (c.323G>A)

2733W
A de novo deletion at 16q24.3 involving ANKRD11 in a Japanese patient with KGB syndrome. S. Miyake1, A. Murakami1, N. Okamoto1, M. Saka-moto1, H. Satitsu1, N. Miyake2, N. Matsutomo1. 1) Department of Human Genetics, Yokohama City University, Yokohama, Japan; 2) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 3) Takarazuka Municipal Center for Handicapped Children, Takarazuka, Japan.

KGB syndrome is a rare autosomal dominant congenital syndrome comprising developmental delay with various neurological involvements, macrodactyly of the upper central incisors, characteristic facial dysmorphism, and skeletal anomalies. ANKRD11 has recently identified as the gene responsible for this syndrome. To date, there have been only several KGB syndrome families described, each carrying a single base substitution or a small deletion of this gene. Here we present a Japanese patient with clinically confirmed KGB syndrome carrying a de novo deletion at 16q24.3 disrupting ANKRD11. He had characteristic facial appearance, macrodactyly of the upper central incisors, hand anomalies, delayed bone age and intellectual impairment without autistic features. Hypoplasia of the cerebellar vermis was pointed out by brain Magnetic Resonance Imaging. Copy number analysis using an Affymetrix Cytogenetics Whole-Genome 2.7M Array and Quantitative realtime PCR on the patient and his parents revealed that he had a de novo 690-kb deletion at 16q24.3 involving part of ANKRD11. Interestingly, the deleted region overlaps with the critical region for 16q24.3 microdeletion syndrome, featuring autosomal dominant intellectual impairment and autism spectrum disorder. We compared the clinical aspects of KGB syndrome to those of 16q24.3 microdeletion syndrome. Although dental information is scarce among the patients with 16q24.3 microdeletion syndrome, the other clinical features, such as macrodactyly of central incisors, are quite similar. However neurological involvement seemed to be different. In 16q24.3 microdeletion syndromes, intellectual impairment is relatively mild and autistic spectrum disorders are frequently observed , while in KGB syndrome, intellectual disability range from mild to severe, and common behavioral disturbances are hyperactivity, attention deficit or easy frustration rather than autistic features. Our patient was very significant in that he had 16q24.3 microdeletion, but his neurological symptoms are more close to those of KGB syndrome. The other spectrum disorder of our patient was hypoplasia of the cerebellar vermis, which had been reported as a rare complication of KGB syndrome, and not of 16q24.3 microdeletion syndrome. Our patient is significant for considering whether these two syndromes are different. It is necessary to study further patients with these two syndromes to clarify this issue.

2734T
Molecular investigations of Polish patients with Beckwith-Wiedemann syndrome. D. Jurkiewicz1, M. Kugaudo1, A. Tariska1, E. Ciara1, D. Piekutowska-Abramczuk1, M. Pelo1, S. Luzzack1, J. Trubicka1, M. Borucka-Mankiewicz1, P. Kowalski1, A. Jezela-Stanek1, A. Ciesliowski1, K. Chraznowska1, M. Krajewska-Walasek1. 1) Department of Medical Genetics, Children’s Memorial Health Institute, Warsaw, Poland; 2) Department of Genetics of Child and Adolescent Psychiatry, Medical University of Warsaw, Warsaw, Poland.

Beckwith-Wiedemann syndrome (BWS) is characterized by overgrowth, macroglossia, abdominal wall defects and a high risk of childhood tumors. BWS is caused by various 11p15 genomic imbalances leading to defective expression of imprinted genes. The genes in 11p15 region are organized into two imprinted domains controlled by two Imprinting Centers: IC1 (H19DMR) and IC2 (KvDMR). Molecular defects of the region resulting in BWS phenotype include loss of methylation at IC2 (50%), paternal UPD of 11p15 (20%), gain of methylation at IC1 (5%) mutations in CDKN1C gene (5%), and chromosomal rearrangements (2%). In ~20% of the patients a molecular alteration is unknown. Forty-nine patients with clinical symptoms of Beckwith-Wiedemann syndrome were investigated. The molecular analysis was performed by methylation sensitive multiplex ligation-dependent probe amplification (MS-MLPA) on leukocyte DNA. Hypomethylation in IC1 region was found in 23 patients including 2 monoyzotic twins. Two patients had hypermethylation in IC2 region. Paternal UPD was identified in 8 patients and was subsequently verified by microsatellite analysis. Two patients carried duplications involving both IC1 and IC2 regions, that were confirmed by arrayCGH. Overall the study revealed the presence of mutations in 35 patients (71%). The molecular background of ~30% of studied cases is still unknown. Of this group of patients the presence of mutations in the CDKN1C gene is possible (the analysis is under way). Moreover, failure to detect UPD in leukocytes due to somatic mosaicism associated with this etiology can not be excluded. There is also a possibility of the presence of molecular alterations in other genomic loci. The presented study is a first complex molecular characterization of a significant group of Polish BWS patients. The pattern of identified genetic defects is comparable with other western populations. The study was financed by National Science Centre, project no. 1149/B/P01/2011/40 (NN407119490) and EU Structural Funds, project POIG.02.01.00-14-059/09.
2735F
Gynecologic Issues in patients with Smith-Lemli-Opitz Syndrome. M.A. Mendez1, S.K. Conley2, F.D. Porter2. 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD; 2) Program in Developmental Endocrinology and Genetics, NICHD, NIH, Bethesda, MD.

Objective: To evaluate the gynecologic issues in females with Smith-Lemli-Opitz Syndrome Background: Smith-Lemli-Opitz Syndrome (SLOS) is a rare autosomal recessive multiple anomaly syndrome of abnormal cholesterol metabolism that is caused by mutations in 7-dehydrocholesterol reductase (DHCR7). Though the clinical spectrum is varied, manifestations include microcephaly, growth retardation, moderate to severe intellectual disability, postaxial polydactyly, characteristic facial features. Current treatments include cholesterol supplementation, early intervention referral and physical/occupational/speech therapy. Despite an expanded phenotypic description of patients with SLOS, there is little information available about the gynecologic issues in this patient population. Methods: Seven females with SLOS were evaluated at the National Institutes of Health Clinical Center. Testing included gynecologic history, laboratory testing, review of outside records, physical examination and pelvic imaging if indicated. Results: The patients ranged in age from 3-36 years. Six patients had normal external genitalia; 1 had redundant perineal tissue. One patient had premature pubic hair development at age 3.5 years. Median age at menarche was 12.5 years (range 10-14 years). Four patients had irregular menstrual cycles and 5 had significant premenstrual mood swings/behavioral problems. Four patients were on hormones to regulate their cycles or help with premenstrual symptoms. One patient had to discontinue hormone use due to elevated liver transaminase levels. Two patients underwent hysterectomy due to difficulties managing menstrual periods. Dysmenorrhea was reported in 3 patients and was attributed by parents to premenstrual and recurrent vaginal infections. No pregnancies were reported. Four patients underwent pelvic imaging with normal findings in all patients. Conclusions: There is a paucity of information in the literature about gynecologic issues in females with SLOS; families and physicians often have questions about puberty, management of menstrual periods and premenstrual behavioral issues. Questions also arise regarding vulnerability in this patient population. Gaining a better understanding of the range of gynecologic problems in SLOS will expand the phenotype and is the first step toward developing gynecologic therapeutic strategies for this patient population.

2736F
Maternal Uniparental Disomy 16 in an infant with intrauterine growth retardation, dysmorphic features, multiple congenital anomalies and dermatoglyphics features suggestive of chromosomal abnormalities: A neglected consideration. Y. Lacassie1, M. Narayanay2. 1) Dept Pedi/ Div Clin Gen, LSU Hth Sci Ctr and Children's Hospital, New Orleans, LA; 2) LSU School of Medicine in New Orleans.

At the 47th Annual Meeting of the ASHG in 1997, the first author reported that the absence of ‘d’ triradius predicted a 50% chance of a chromosomal abnormality. At the 59th meeting in Hawaii in 2009, he presented how dermatoglyphics in newborns and fetuses can indicate cryptic chromosomal disorders, including microdeletions and duplications. We present a 3-month-old male with history of intrauterine growth restriction born to a 38yo mother at 36+6 WGA weighting 1,342g. On day 2 of life, the patient experienced severe respiratory distress. After appropriate intervention and stabilization, a complex congenital heart defect (CHD) and bilateral pulmonary hypoplasia were diagnosed and various dysmorphic features were noticed. Abnormally high levels of plasma amino acids, urine organic acids, and carnitine were found; however, the newborn screen was negative. Initial chromosomal analysis (450-500 bands resolution level) and aCGH (180,000 probes, 300kb resolution) were normal. The patient was transferred to Children’s Hospital New Orleans at 53 days of life. Further work up showed multiple cardiovascular anomalies, including mitral insufficiency, VSD, PFO, left pulmonary artery stenosis, non-compactd LV, hypoplastic aorta, and left SVC. Also, posteriorly rotated ears, antverted nostrils, first degree hypoplasia, cryptorchidism, right inguinal hernia, severe growth retardation, the presence of vertical crease in the soles secondary to gap fourth toe, secondary to an unstable mutation of the DMPK gene locus 19q13.3. Precocious puberty was evident at birth. Physical examination revealed a high risk of falling in 34.78% of the evaluated patients. Conclusions: To our knowledge, this is the first report of maternal UPD 16 demonstrating the maternal origin of UPD. A literature review of 54 cases reveals that the major clinical features of patients with maternal UPD 16 are highly variable. They may include moderate to severe IUGR, CHD, pulmonary hypoplasia, inguinal hernia, TEF, anal atresia, and dysmorphic facies. This patient shows that UPD 16 should be considered in the differential diagnosis of apparent chromosomal microdeletions or duplications as they may present similar phenotypic findings, including dermatoglyphics.

2737T

Introduction: Myotonic dystrophy type 1 (DM1) is a progressive and multisystem degenerative neuromuscular disease, characterized by myotony and muscular weakness, secondary to an unstable mutation of CTG repeats in the DMPK gene locus 19q13.3. Precocious pubescent has been reported in DM1 patients indicating a compromised auditory system. The present work shows the results from audiological and vestibular evaluation in Mexican patients with DM1 from the National Rehabilitation Institute (INR). Methods: 42 patients with molecular diagnosis for DM1 were analyzed. An extensive audiological evaluation was performed. Audiograms classified the range, nature and degree of hearing loss and speech, detected the sensorial or conductive nature of hearing loss. Tympanometry, acoustic reflex threshold, and transient otoacoustic emissions were part of the evaluation. The impact of the vestibular disease in daily life was evaluated by the Dizziness Handicap Inventory (DHI). Results: In 84 ears analyzed, 26 showed normal hearing (30.95%), 38 presented several degree of affection (69.05%), being the most prominent effect the mild-sensory hearing loss on low frequencies (22.62%) and mild sensorial hearing loss on high frequencies (19.05%). Most of the patients showed auditory alteration in both ears. High frequency audiometry revealed falls from 40 to 50 db in 57% of the ears. The most prevalent tinnitus curve was of type A and A*, with absence of acoustic reflex thresholds in the 60% of them. Transient otoacoustic emission revealed an inadequate response in 60% of ears. The emotional scale of vestibular evaluation indicated no disability in the 78.26% of the patients and moderate disability in 21.73%. Functional scale showed 21.73% and 8.69% of moderate and severe disability respectively. Physycal scale evaluation showed a 13.04% of moderate and severe disability respectively. Finally the DGI revealed a high risk of falling in 34.78% of the evaluated patients. Conclusions: DM1 can be associated to sensorial and vestibular affection of the base and apex of the cochlea where the outer hair cells play a key role in sound perception through a contractile-like movement. Future research efforts focused on the cells in the context of CTG repeats should be done to understand the pathophysiology of the disease at this level.

2738F

Hearing impairment (HI) affects 1 in 650 newborns which makes it the most frequent congenital sensory impairment. Nonsyndromic HI is very common and the presence of 37 different genes with a wide variety of functions have been identified. Mutations in GJB2 are the most common genetic etiology of prelingual non-syndromic sensorineural hearing loss. Rikkert et al reported in 2005 a total of 83 different mutations in GJB2; they found that 35deIG mutation was the most common with an allelic frequency of 72.44%; they also found a reduced allelic frequency of 0.3% for the S199F mutation. In the study of Olarte M et al in Colombian population they reported that the most common mutations were 35deIG and S199F; the allelic frequency for S199F mutation was 42.1%. In this paper we report a case of a Colombian 8 years old female child with a history of bilateral sensorineural hearing loss diagnosed at 5 years old. Her parents were non-consanguineous and her family history was unremarkable. She was born at term after a normal pregnancy. Mother denied exposure to teratogens or infections during pregnancy. In order to establish the etiology of her sensorineural hearing loss we sequenced the gene GJB2 and GJB6 was performed, the report showed S199F homozygous mutation in GJB2 and a normal sequence in GJB6. Most patients with mutations in GJB2 present a prelingual hearing loss but in this case our patient has a S199F mutation in GJB2 with a non classical postlingual bilateral sensorineural hearing loss and an atypical audiometry pattern.
2739W

Simpson-Golabi-Behmel syndrome is a rare overgrowth syndrome caused by the GPC3 mutation at Xq26 and is clinically characterized by multiple congenital abnormalities, mental retardation, pre/postnatal overgrowth, distinctive craniofacial features, macrocephaly, and organomegaly. Although this syndrome is known to be associated with a risk for embryonal tumors, similar to other overgrowth syndromes, the pathogenetic basis of this mode of tumorigenesis remains largely unknown. Here, we report a boy with Simpson-Golabi-Behmel syndrome who had a germline loss-of-function mutation in GPC3. At 9 months of age, he developed hepatoblastoma. A comparison of exome analysis results for the germline genome and for the tumor genome revealed a somatic mutation, p.Ile35Ser, within the degrada-
tion targeting box of beta-catenin. The same somatic mutation in CTNNB1 tumor genome revealed a somatic mutation, p.Ile35Ser, within the degrada-
tion targeting box of beta-catenin. The same somatic mutation in CTNNB1

2740T
Establishment and validation of iPSC cells and knockout mice for der-matan 4-O-sulfotransferase 1 (D4ST1)-deficient Ehlers-Danlos syndrome (EDDS). J.C Prieto1,2, G. Giraldo2, S.Mizumoto3, M.Kobayashi1,4, F. Fukushima1, S. Saka1, 1) Department of Molecular Therapy, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan; 3) Department of Histology and Embryology, Shinshu University School of Medicine, Matsumoto, Japan; 4) Scleroprotein and Leather Research Institute, Tokyo University of Agriculture and Technology, Faculty of Agriculture, Fuchu, Japan; 5) Laboratory of Proteoglycan Signaling and Therapeutics, Hokkaido University Graduate School of Life Science, Sapporo, Japan; 6) Miya Kobayashi: Department of Food and Nutritional Environment, College of Human Life and Environment, Kinjo Gakuin University, Nagoya, Japan; 7) Department of Molecular Pathol-
ogy, Shinshu University Graduate School of Medicine, Matsumoto, Japan; 8) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 9) Department of Cell Modulation, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto, Japan; 10) Department of Dermatology, Dokkyo Medical University, School of Medicine, Mibu, Japan.

D4ST1-deficient Ehlers-Danlos syndrome (EDDS) was characterized by progressive multisystem fragility-related complications. The first case of D4ST1-/- mice was demonstrated, suggesting these mice to reflect glycobiological abnormalities of Chst14-/- mice. Complete loss of DS in patients' urine was also detected. Because the patients suffer from progressive multisystem fragility-related complications, appropriate disease modeling is indispensable in view of developing etiology-based therapy. In this study, we report establishment and validation of induced pluripotent stem cells (iPSCs) and knockout (Chst14-/-) mice. Cultured skin fibroblasts were obtained from a patient and transduced into iPSCs. Morphologically, patient-derived iPSCs (P-iPSCs) had a smaller size and more vacuoles than control iPSCs (C-iPSCs) derived from healthy individual. Apoptosis, undifferentia-
tion status, or pluripotency were not different between these iPSCs. Decorin staining on teratoma samples derived from P-iPSCs using human decorin antibody was weaker than that from C-iPSCs, similar to the results from patients' skin specimens. Neural progenitor cells and neurons were differenti-
tiated, and neural differentiation potency was weaker in P-iPSCs than that in P-iPSCs. Frozen sperm from Chst14-/- male mice were obtained from the Mutant Mouse Regional Resource Center. After reproduc-
ing Chst14-/- mice, Chst14-/- mice were generated and preliminary phenotypic studies were undertaken. Complete loss of DS in urine from Chst14-/- mice was demonstrated, suggesting these mice to reflect glycobiological abnormalities in DDEDS. Although postnatal weight gain was smaller in Chst14-/- mice than in Chst14+/- or WT mice, no apparent congenital abnormalities or multisystem fragility-related manifestations were noted. Further studies to delineate pathophysiological features of cells from other systems differenti-
tiated from iPSCs and longitudinal physical features including pathological abnormalities of Chst14-/- mice are necessary.

2741F
Report of a Colombian case of Werner mesomelic syndrome with eight fingers in both hands and feet. J.C Prieto1,2, G. Giraldo1, T. Pineda1, 1) Inst de Genetica Humana, Univ Javeriana, Bogota, Colombia; 2) Hospital La Victoria, Secretaria Distrital de Salud Bogota, Colombia.

Werner mesomelic syndrome (WMS) is an autosomal dominant disorder first described by Werner in 1912. It is characterized by hypo or aplasia of the tibiae with remarkable short stature associated with preaxial polydactyly of the hands and feet and/or five-fingered hand with absence of thumbs. Wieczorek et al identified the molecular basis of the syndrome in 2010; they established that this condition is caused by a specific point mutation in the ZRS (the sonic hedgehog regulatory region) at position 404. In this paper we report the case of a Colombian 1 year old infant patient with Werner mesomelic syndrome. She was the first daughter of non-consanguineous young parents and her family history was unremarkable; she presented with racongenic and mesomelic shortening of the lower limbs with bilateral hypoplasia and bilateral absence of the halluc; she also presented bilateral triphalangeal thumb and preaxial polydactyly with 8 fingers in both hands and feet. This is the first case of Werner mesomelic syndrome reported in Colombia and we describe new phenotypic findings.
2742W

A Novel WNT7A Mutation Causes Autosomal Recessive Al-Awadi/Raas-Rothschild/Schinzel Phocomelia Syndrome. Y. Sahin1, P.O. Simsek-Kiper2, A. Cetinkaya1, GL. Ulte1, K. Boduroglu1. 1) Department of Medical Genetics, Hacettepe University, Ankara, Turkey; 2) Department of Pediatrics, Department of Pediatric Genetics, Hacettepe University, Ankara, Turkey.

Three generation family with kyphomelic dysplasia suggests autosomal dominant inheritance. This observation has important implications. Even the proband had an identical phenotype. All had a normal cognitive development.

2743T

Three generation family with kyphomelic dysplasia suggests autosomal dominant inheritance. O. Vanakker1, U. Fränkel1, B. Callewaert1, 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Department of Paediatrics, Zorgsaam Hospital, Terneuzen, The Netherlands.

Kyphomelic dysplasia (KD) is a rare skeletal dysplasia belonging to the group of bent bone dysplasias. It represents a heterogeneous group of disorders, with at least three distinct entities identified: Schwartz–Jampel syndrome (SJS), Cartilage Hair Hypoplasia (CHH) and ‘true’ KD. For the latter, the genetic background is unknown and autosomal recessive inheritance is presumed based on several sporadic cases with parental consanguinity. We report a three-generation family presenting with a true KD phenotype.

The proband was born at term after an uneventful pregnancy. Her length and weight at birth were 37cm and 1578g respectively. She had severe rhizomelic shortening of upper and lower limbs and radiographs showed stubby femora which were extremely curved, iliac wings were short and widened; the humeri were also short and dumbbell-shaped. The mother of the proband has a similar disproportionate short stature phenotype. Radiographs taken at birth revealed severe femoral bowing, though less prominent compared to her daughter. Follow-up radiographies in the mother demonstrated a gradual decrease of the bowing over several years. The maternal grandfather of the proband had an identical phenotype. All had a normal cognitive development.

2744F

Further evidence of brain anomalies related to ALX4 mutations: possible genotype-phenotype correlation. T. Almeida1, M. Valente1, G.L. Yamamoto1, C.A. Kim1, M.R. Passos-Bueno2, D.R. Bertola1, 1) Genetic Unity, Children’s Institute, São Paulo, São Paulo, Brazil; 2) Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, São Paulo, Brazil.

Brain anomalies has been found in patients with loss-of-function mutations in ALX4, either in the homozygous state, with the frontonasal phenotype, showing abnormalities in corpus callosum and cerebellar hypoplasia, or in the heterozygous state, with parietal foramina, presenting polymicrogyria, high insertion of the tentorium cerebri, bilateral choroid plexus cysts, corpus callosum agenesis and microcephaly. We describe a cranial MRI of four cases with ALX4 mutations. Patient 1, with a homozygous missense mutation in the homeodomain (p.R215W) and a frontonasal phenotype, showed occipital and cerebellar hypoplasia, high insertion of the tentorium cerebri, agenesis of the septum pelucidum and occipital simplified gyra, with no abnormalities in corpus callosum, previously described for all patients with the frontonasal phenotype. Patient 2, heterozygous for the same mutation, showed no brain anomalies. Patients 3 and 4 (mother and son) with a heterozygous frameshift mutation after the homeodomain (p.D296fs+21) and a frontonasal phenotype, showed no structural anomalies. Contrary to what is observed in the literature for the frontonasal patients, our study describes two individuals with this phenotype without brain anomalies.

2745W

Further occurrence of ‘TMCO1 defect syndrome’ in a non-Amish population: the first South American patient presenting the same p.R87X mutation recently found in a Turkish individual. W.A.R. Baratela1, G.L. Yamamoto1, M.R. Passos-Bueno2, D.R. Bertola1, 1) Unidade de Genética, Instituto da Criança - Hospital das Clínicas - Universidade de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Centro de Estudos do Genoma Humano - Departamento de Genética e Biologia Evolutiva - Instituto de Biociências - Universidade de Sao Paulo.

Autosomal recessive syndrome of craniofacial dysmorphism (brachycephaly, flat face, arched eyebrows, synophrys, ocular hypertelorism, wide nose bridge, high-arched palate, cleft lip and palate, microdontia of primary teeth, gingival hyperplasia and low-set ears), skeletal anomalies (Sprengel deformity, plaeus excavatum, rib anomalies, vertebral fusions and club foot), and developmental delay/mental retardation associated with TMCO1 defects was first diagnosed in 11 individuals, all from the Old Order Amish of northeastern Ohio. TMCO1 homozygous nonsense mutation in the coiled coil region of the protein was found in all 9 tested patients, with a carrier frequency in the community of 0.7%. Recently a similar case from Turkey was published, with a different mutation in TMCO1, but in the same protein region, with a rather more profound developmental impairment. We describe a 28 yo lady, daughter of a first cousin Brazilian couple, with normal brain malformations. She presented with tall stature, bushy and arched eyebrows, synophrys, ocular hypertelorism, divergent strabismus, clift lip and palate and gingival hypertrophy. She also had polydactyly, presented in other members of the family in an autosomal dominant pattern. She walked at 2y4mo, first words with 1y6m, attended special education from the beginning. At 3yo developed seizures and started with antiepileptic drugs. Skeletal survey showed rib fusions and scoliosis. Echocardiogram, and abdominal ultrasound were normal. Molecular studies with exome sequencing found a homozygous mutation in TMCO1, c.C259T (p.R87X), the same reported in Turkey. We describe a further occurrence of ‘TMCO1 defect syndrome’ in a non-Amish population: the first South American patient, indicating that this disorder is probably more widespread. The family of our patient is from a small city in the Northeast of Brazil, in which there was no intermarriage. The mutation is not one of the 6 mutations previously linked to Al-Awadi/Raas-Rothschild/Schinzel Phocomelia Syndrome.
2746T
Sibs with hydrops fetalis, arthrogryposis multiplex congenital, neural migration disorder, adrenal and pulmonary hypoplasia and renal abnormalities associated with a mutation in the FAT4 gene, encoding a giant cadherin. D. Chitayat1,2, T. Uster3, P. Shannon4, M. Strour1, S. Robertson5, J. Michaud1,1. 1) Prenatal Diag & Med Gen, Mount Sinai Hosp, Toronto, ON, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Ontario, Canada; 3) Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Ste-Justine Hospital Research Center, Montreal, Quebec, Canada; 5) Department of Pediatrics and Child Health, Dunedin School of Medicine, University of Otago, New Zealand.

The human FAT gene family consists of the FAT1, FAT2, FAT3 and FAT4 genes. Human FAT family genes encode large proteins with extracellular cadherin repeats, EGF like domains, and laminin G like domain(s). FAT4 is involved in the maintenance of planar cell polarity and inhibition of cell proliferation. FAT4 gene has been reported to be mutated in several types of human cancers, such as melanoma, pancreatic cancer, gastric cancer and hepatocellular carcinoma. However, so far no human condition has been reported to be caused by germline mutations in this gene. We report a male and female sibs, born to a consanguineous Turkish couple, with hydrops fetalis (HF), neural migration disorder, joint contractures, pulmonary and adrenal hypoplasia and renal abnormalities. To our best knowledge this is a new, familial recessive condition. The parents were healthy and first cousins of Turkish descent. Patient A, a male, was born at 31.8 weeks gestation following a pregnancy complicated with hydrops fetalis, mild ventriculomegaly, misalignment of the skull bones and lumbar lordosis. He died shortly after delivery and autopsy showed HF, cleft palate, ambiguous genitalia, bilateral talipes equinovarus, campodactyly of the fingers and clenched fists. There was a butterfly T5 vertebral and hypoplastic lungs, kidneys and adrenal glands. The brain showed megalencephaly, abnormal gyration and an unusual pattern of cytoarchitectural disturbance in the cerebral cortex. The karyotype was 46, XY. Patient B, a female, the product of the couple's second pregnancy, was found to have an increased nuchal translucency and body edema on 12 weeks ultrasound. Ultrasound at 20 weeks gestation showed bilateral talipes equinovarus, flexed wrists, echogenic bowel, right pleural effusion and a micrognathia configuratio. The pregnancy was terminated at 21 weeks gestation and the autopsy showed the same findings as in her late brother apart from her female internal and external genitalia. Microarray analysis was normal and female. Extensive analysis on the parents showed no abnormality. Exome sequencing showed a homozygous mutation in the FAT4 gene.

2747F
Genetic heterogeneity in Mabry syndrome: a novel phosphatidylinositol glycan (GPI) anchor deficiency disorder. D. Cole1, T. Roscio2, M. Nezarati3, E. Sweeney2, P.N. Krawitz3, H. van Bokhoven4, C. Marcelis4, B. DeVries5, D. Andrade6, W.M. Burnham7, A. Munnich8, M. Thompson9, 1) Dept Laboratory Medicine, Univ Toronto, Toronto, ON, Canada; 2) Dept of Women's and Children's Health, Sydney Children's Hospital and University of New South Wales; 3) Dept of Genetics, New York General Hospital. Toronto, Ontario, Canada; 4) Royal Liverpool Children's Hospital, Liverpool, UK; 5) Inst fur Medizinische Genetik, Charité Universitätsmedizin Berlin, Germany; 6) Dept of Human Genetics, University Medical Centre St. Radboud, Nijmegen; 7) Division of Neurology, Toronto Western Hospital, Toronto, Canada; 8) Department of Pharmacology, University of Toronto; 9) INSERM U781-Université Paris Descartes-Hôpital Necker-Enfants Malades, Paris.

Mabry syndrome (hyperphosphatasia with developmental disability) was first described in 1970 (OMIM#239300). At first considered rare, improved sequencing showed a homozygous mutation in the FAT4 gene. De novo mutations in the FAT4 gene are associated with abnormalities of the phosphatidylinositol glycan (GPI) anchor disruption. Not all patients with a clinical diagnosis of hyperphosphatasia with neurologic deficit result from known GPI disruptions. We present data on patients that suggest there are further subtypes of the disorder.

2748W
Exome sequencing of the X-chromosome in Aicardi syndrome. C. Lund, H.S. Sorte1, Y. Sheng2, M.D. Vigeland3, G. Rabsy4, K.K. Selmer1. 1) Department of Ophthalmology, University of Iowa, Iowa City, IA; 2) Department of Pediatrics, University of Iowa, Iowa City, IA; 3) X-ray Crystallography Facility, University of Iowa, Iowa City, IA; 4) Center for Integrative Brain Research, Seattle Children’s Research Institute, Seattle, WA; 5) Department of Neurology, Northwestern University Evanston, IL; 6) Department of Computer Science, University of Iowa, Iowa City, IA; 7) Department of Biology, University of Iowa Iowa City, IA; 8) Aichi-Gakuin University, Japan; 9) Odonto Maxillo Facial Hospital, Vietnam; 10) IRCCS Casa Sollievo della Sofferenza, Italy; 11) Department of Experimental Medicine, Sapienza University, Rome, Italy; 12) Department of Medicine and Surgery, University of Salerno Salerno, Italy; 13) Diwanchand Satyapal Imaging Research Centre, India.

Background The Dandy-Walker spectrum of disorders including autosomal dominant Dandy-Walker malformation and occipital cephaloceles. (ADDWOC) are characterized by variable cerebellar hypoplasia, meningeal anomalies, and occipital skull defects. We have previously reported deletions of the X-chromosome of seven patients, we were not able to reveal the etiology of Aicardi syndrome. Assuming that the hypothesis of X-linked dominant inheritance is correct, the results of this study suggest that the causal locus lies in a non-coding region or in a gene poorly covered by the exome sequencing kit. However, further studies of these patients will involve trio testing and examination of the autosomal chromosomes.

2749T
Mutations in extracellular matrix genes NID1 and LAMC1 cause autosomal dominant Dandy-Walker malformation and occipital cephaloceles. V. Mahajan1, B. Darbro2, L. Gakhar3, J. Skeie4, E. Campbell5, S. Wu6, X. Bing7, K. Milianen8, W. Dobbs9, J. Kessler6, A. Jaiswal10, A. Segre11, J. Manak12, K. Aidding9, S. Suzuki10, N. Natsume10, M. Ono10, H. Dali Hari10, L. Thi Viet8, S. Loddo6, E. Valente10,12, L. Bernardini6, P. Ghonge13, P. Ferguson14, A. Bassuk1. 1) Department of Ophthalmology, University of Iowa, Iowa City, IA; 2) Department of Pediatrics, University of Iowa, Iowa City, IA; 3) X-ray Crystallography Facility, University of Iowa, Iowa City, IA; 4) Center for Integrative Brain Research, Seattle Children’s Research Institute, Seattle, WA; 5) Department of Neurology, Northwestern University Evanston, IL; 6) Department of Computer Science, University of Iowa, Iowa City, IA; 7) Department of Biology, University of Iowa Iowa City, IA; 8) Aichi-Gakuin University, Japan; 9) Odonto Maxillo Facial Hospital, Vietnam; 10) IRCCS Casa Sollievo della Sofferenza, Italy; 11) Department of Experimental Medicine, Sapienza University, Rome, Italy; 12) Department of Medicine and Surgery, University of Salerno Salerno, Italy; 13) Diwanchand Satyapal Imaging Research Centre, India.

Background The Dandy-Walker spectrum of disorders including autosomal dominant Dandy-Walker malformation and occipital cephaloceles (ADDWOC) are characterized by variable cerebellar hypoplasia, meningeal anomalies, and occipital skull defects. We have previously reported deletions of the X-chromosome of seven patients with Aicardi syndrome. Assuming that the hypothesis of X-linked dominant inheritance is correct, the results of this study suggest that the causal locus lies in a non-coding region or in a gene poorly covered by the exome sequencing kit. However, further studies of these patients will involve trio testing and examination of the autosomal chromosomes.
2750F
Mapping a new locus for autosomal dominant nonsyndromic comitant strabismus. M. Patel1,2, S. Ye1,2, C. Shyr3, Z. Zong1,2, M. Thomas1, P. Power1, N. Roslin1, S. Narasimhan1, D. Giaschi2, W. Wasserman1,2,3.
1) Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics; 3) Child and Family Research Institute; 4) FORGE Canada Consortium; 5) Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, BC, Canada.

Strabismus, defined as a lack of parallel gaze, is a common problem found at an elevated frequency in a large number of different syndromes, precluding the use of syndromology to unravel the genetic basis of this common disorder. The lack of common molecular or cell biological features between all the syndromes featuring strabismus may also suggest that there are many possible disruptions that can result in an interruption of parallel gaze. One way to address this complexity is to study families with Mendelian inheritance of the trait of interest. We report here a 7 generation pedigree with many affected family members in an autosomal dominant pattern. Affected members had isolated esotropic or hypertropic comitant strabismus that was nearly always congenital and of varying severity. No other ocular or neurological abnormalities segregated in the pedigree and there were no minor anomalies that segregated with the trait. A high density genome-wide mapping study was undertaken in parallel with exome sequencing. Results will be reported at the meeting. Study of this rare family provides an opportunity to uncover a molecular entry point into the biology of nonsyndromic strabismus.

2751W
Exome sequencing identifies mutations in a gene not previously related to skeletal dysplasias in patients with spondylometaphyseal dysplasia. G.L. Yamamoto1,2, X. Ye1,2, C. Shyr3, Z. Zong1,2, M. Thomas1, P. Power1, N. Roslin1, S. Narasimhan1, D. Giaschi2, W. Wasserman1,2,3.
1) Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics; 3) Child and Family Research Institute; 4) FORGE Canada Consortium; 5) Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, BC, Canada.

Exome sequencing. Results will be reported at the meeting. Study of this rare family provides an opportunity to uncover a molecular entry point into the biology of nonsyndromic strabismus.

2752T
A novel missense mutation in PRPS1 leads to PRS-I deficiency in females displaying retinitis pigmentosa and variable expression of a neurologic phenotype. B. Almqvist1,2, J. Liang2, P. Fernandez3, M. Con- tor4, Y. Guo5, B. Keating1,4,6, J. Zhang7, H. Hakonarson1,4,5, X. Xu2, C. Ayuso8.
1) Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) Department of Genetics and Genomics, IIS-Fundacion Jimenez Diaz Hospital, 28040, Madrid, Spain; 4) Division of Human Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA; 5) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

Phosphoribosyl pyrophosphate synthetase I (PRPS1) is a key enzyme in nucleotide metabolism. Missense mutations in PRPS1 may lead to PRS-I deficiency and cause three different phenotypes: Arts Syndrome (MIM 301835), linked Charcot-Marie-Tooth (CMTX5, MIM 311707) or X-linked non-syndromic sensorineural deafness (DFN2, MIM 304500). All three have in common to be rare, X-linked recessively inherited and to display variable levels of central and peripheral neuropa thy in males affected. In the current study, we have identified a novel missense mutation in PRPS1 leading to PRS-I deficiency in 4 affected females across 3 generations of a family with a phenotype consisting of retinitis pigmentosa (RP) and variable expression of optic atrophy, ataxia, peripheral neuropathy and hearing loss. Four members from the family were subjected to whole exome sequencing. After variant filtering analysis assuming dominant inheritance, a novel missense variant in PRPS1 was selected among 107 nonsynonymous variants segregating in the family. Sanger sequencing in 12 members of the family showed complete segregation of the mutation with the disease. PRS-I enzymatic activity in erythrocytes from three affecteds evidenced variable levels of deficiency, with severity correlating with the age of onset of ophthalmologic manifestations. Further assessment of X inactivation skewing confirmed different patterns of X inactivation in the four affecteds, which may be also responsible for the variable phenotypic expression in the females. However, since no male was found affected, a lethal effect of the mutation in males cannot be ruled out. In addition to the observation of PRS-I deficiency in females, this is the first time that RP is part of the myriad of symptoms displayed by this family the first and only common manifestation to all affecteds. Visual cycle is especially sensitive to GTP levels, which are expected to be low in the family, and that may lead to an eventual death of photoreceptor cells and would explain the RP phenotype. A functional study and elucidation of the mechanism underlying RP. This study also highlights the enormous phenotypic variability observed in PRS-I related syndromes inter and intra family.

2753F
1) Genetics Dept, Varamin-Pishva Branch, Islamic Azad University, Varamin, Iran, varamin, Iran; 2) Department of Medical Genetics, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran; 3) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Androgen receptor gene mutations are the leading cause of sex development disorders exhibited by sexual ambiguity or sex reversal. In this study, seven families containing patients clinically diagnosed as androgen insensitivity syndrome (AIS) were genetically evaluated through cytogenetic and molecular analysis including karyotype, polymerase chain reaction (PCR) and sequencing of SRY and AR genes. Two brothers and their mother were hemizygous and heterozygous respectively for c.2522G>A variant, while one of their healthy brother was completely normal hemizygous. Family 2 assessment demonstrated the c.629G>A (rs6152) variant in two siblings who were reared as girls. The SRY gene was intact in all of the study’s participants. We could not find any mutation or polymorphism in neither AR nor SRY genes of the other five patients from five families. Our findings indicated that AR mutations in development of morbidities such as disabling sex life and in particular prostate cancer in AIS patients, definitive diagnosis based on molecular genetic approaches accompanied by comprehensive genetic counseling could have promising impacts in clinical management and also in prenatal diagnosis of prospect offspring.
Ocular Findings in the Marfan Syndrome. J.H. Maumenee1, S. Wehri2, WW. Xu3, S. Rahman4, S. Kurup5, I. Kassem1, M.K. Durbin6, N. Azar7, A.A. Fawzi8, A. Lyon4, M.B. Meis1, Chicago Marfan Eye Consortium. 1) Ophthalmology, University of Illinois Eye and Ear Infirmary, Chicago, IL; 2) Lurie Children’s Hospital of Chicago, Chicago, IL.

Purpose: To identify diagnostic criteria and risk factors for vision loss in Marfan patients based upon clinical examination, biomicroscopy and OCT imaging of the posterior segment. Material and Methods: 73 patients were seen at Children’s Hospital Ophthalmology Clinic in Chicago during the 2012 National Marfan Foundation Conference. Fifty-six patients fit diagnostic criteria for Marfan syndrome and underwent extensive imaging using Humphrey visual fields, fundus photography, slit lamp, indirect ophthalmoscopy, biomicroscopy, OCT measurements using the Macular Cube 512x128 and Optic Disc Cube 200x200 of the Cirrus HD-OCT (Carl Zeiss Meditec, Dublin, CA). Criterion for OCT inclusion was signal strength of ≥ 8 out of 10 without artifacts; the data were analyzed using Cirrus software. Retinal examinations were performed by a vitreoretinal specialist. Color ultra-widefield retinal images were obtained using the Optos 200Tx on 54 patients (108 eyes). The ultra-widefield images were evaluated by a masked vitreoretinal specialist. The color composite images were analyzed with Optos imaging review software. Results were analyzed using Pearson’s correlation, 2-tailed tests or one sample t-test. Results: Mean age of patients was 20.7 ± 7 years (range 2 to 56). Keratometry obtained on 54 eyes showed flattening of the cornea (mean K=41.50D; SD=2.08). Corneal thickness, (mean=542.2 μm; SD=42.25); deepened anterior chamber, (mean=3.6mm; SD=0.61), reduced lens diameter (mean=3.84mm; SD=0.48); and increased axial length, (mean=25.39mm; SD=1.91). A dislocated lens was observed in 32% of patients. OCT data showed that 27% of eyes had RNFL thickness <5% suggestive of RNFL loss; 24% had peripheral retinal pathology significantly to the diagnosis of Marfan syndrome. 27% of eyes had RNFL thickness in 32% of patients. OCT data showed that 27% of eyes had RNFL thickness with 24% of eyes analyzed using Optos 200 TX. Corneal curvature, AC depth, lens thickness and axial length of the globe (p<.05) can contribute significantly to the diagnosis of Marfan syndrome. 27% of eyes had RNFL thickness <5% suggestive of RNFL loss; 24% had peripheral retinal pathology, both indicative of need for close patient observation.
The association of Coats disease with intrathecal growth retardation, intracranial calcification, leukodystrophy, brain cysts, osteopenia, and gastrointestinal bleeding defines Coats plus syndrome caused by mutations in the CTC1 gene, encoding conserved telomere maintenance component. Here, we report on a child with exudative retinopathy, cerebral calcifications, duodenal atresia, preaxial polydactyly, microopenis, microcephaly and short stature: a new syndrome. The identification of a missense mutation in the MBTPS2 gene as the cause of X-linked form of Olmsted syndrome suggests telomere biology and pathology of Olmsted syndrome and has applications in genetic counseling.
2761T
Psychological and cognitive profile in four new patients with MOMO syndrome. C. Passalaquca, M. Garcia, L. Dueñas. Hospital Carlos Van Buren, Valparaiso, Valparaiso, Chile.

MOMO is an acronym for macrosomia, obesity, macrocephaly and ocular anomalies (OMIM 157980). The first patients were presented in 1993 by Moretti-Ferreira et al. with a total of seven patients published thus far. All the cases presented mental disability and in one case autism was described. We present four new cases with a neuropsychological profile, there were referral to geneticist and then to psychologist and child psychiatry. All parent patients signed a consent form to participate in the study. All patients were seen by child psychologist who performed WISC III (adapted for Chilean population) and also with child psychologist who did the MINI-KID test, used to assess the presence of 24 DSM-IV child and adolescent psychiatric disorders as well as the risk of suicide. The patients are three male (11, 12 and 17 years old) and a female of 13 years old. The four patients fulfill the criteria of macrocephaly, obesity and ophthalmological abnormalities; macrosomia at birth was seen in one patient (male, birth weight: 4895 g >>P99) and another had tall stature to date (male, 11 years, height 159 cm +1.8SD). The karyotype and the Prader Willi methylation test were normal in three patients, in the remain patient are still in process. The four patients had an history of developmental delay, after the psychological evaluation two had ranked as intellectual disability (IQ 51 and 58) and the others as borderline intellectual functioning (IQ 77 and 72). The subscales showed in two patients asymmetry between verbal skills and execution functions, in three patients the lower skills were assimilate information and cubes assembly. The MINI-KID demonstrate that three had psychiatric comorbidities (anxiety, phobia and behavior disorder conduct) the other patient (17 years) had borderline intelligent functioning, without psychiatric comorbidities and he is more independent in the living day activities, Patient 3 and 4 are siblings from a healthy non consanguineous couple, Moretti-Ferreira suggested an autosomal dominant inheritance, while Di Donato theorized an autosomal recessive trait. We think that are still few patients to propose an inherence trait, but the existence of brothers with healthy parents is more suggestive to an autosomal recessive inheritance. This is the first neurocognitive evaluation of MOMO patients and we agree with Di Donato who proposed change the M from Macrosomia to Mental (intelectual) disability.

2762F

ARPKD/CHF is characterized by dilated renal collecting ducts resulting in renal insufficiency and CHF often complicated by portal hypertension. It is caused by mutations in PKHD1, which encodes fibrocytin, a protein on the primary cilia-basal body/centriole. The consensus clinical diagnostic criteria for ARPKD/CHF require characteristic kidney and liver involvement, family history consistent with autosomal recessive inheritance, and absence of congenital anomalies. Little is known about pregnancy outcomes for patients with ARPKD/CHF. Patients with other causes of chronic kidney disease are at risk of pregnancy-related worsening of renal function and severity of portal hypertension. Appropriate preconception counseling is critical to inform patients of pregnancy risks and alternative reproductive options.

2763W
X-linked Joubert Syndrome: neuroimaging and clinical features associated with a novel mutation in OFD1. B. Hashemi1, H.M. Branson2, M. Moharram1, G. Yoon1, 2 1) Division of Clinical Genetics, Hospital for sick children, Toronto, Ontario, Canada; 2) Division of Pediatric Neuroradiology, Hospital for sick children, Toronto, Ontario, Canada; 3) Division of Neurology, Hospital for sick children, Toronto, Ontario, Canada.

Oral-facial-digital syndrome type 1 (OFD1; OMIM#311200) is a rare developmental disorder characterized by craniofacial, oral, skeletal abnormalities, cystic kidneys and neurological involvement including mental retardation. The syndrome is caused by mutations in the OFD1 gene and results in male lethality in the first or second trimester. The spectrum of phenotypes associated with mutations in OFD1 continues to expand, and genotype-phenotype correlations are beginning to be recognized. The spectrum includes Simpson-Golabi-Behmel syndrome type 2 (SGB2, OMIM#300209), and X-linked Joubert syndrome (JSB type10) phenotype in addition to the classical OFD1 phenotype with early male lethality. We describe the clinical and neuroimaging findings in a 17 year old male with a clinical diagnosis of Joubert syndrome associated with a novel mutation in the OFD1 gene (JS-10). Our case confirms the reported association of OFD1 gene mutation with X-linked recessive Joubert syndrome. This family is the fifth reported with X linked Joubert syndrome, associated with a novel mutation in exon 17 of the OFD1 gene. We describe unique neuroimaging and clinical features associated with OFD1 mutations and highlight the utility of advanced neuroimaging techniques in characterizing the phenotype.

2764T
Xq11.1-11.2 deletion involving ARHGEF9 in a girl with autism spectrum disorder. G. Bhat1,2, D. LaGrave3, A.N. Lamb4, R. Matafonov1. 1) Pediatrics, New York Medical College, Valhalla, NY; 2) Medical genetics, Pediatrics, University of Texas Medical Branch, Galveston, TX; 3) Department of Cyto-genetics, ARUP laboratories, Salt Lake City, UT; 4) Department of Pathology, University of Utah, Salt Lake City, UT.

We report a de novo ARHGEF9 deletion in a 6 year old girl presenting with autism spectrum disorder, ADHD, severe speech delay, developmental delays and cognitive impairment. Our patient was born at term without any known complications. She walked at 16 months and started speaking at 2 years of age; she is still unable to make two word sentences. Her hearing is normal and she receives speech therapy at school. There is no history of seizures or sensory hyperarousal and she is not dysmorphic. She was diagnosed with ASD and ADHD by a behavior specialist and is currently receiving therapy. Cytogenomic microarray, using the CytoScan HD platform, demonstrated a de novo 82 Kb deletion involving the X chromosome from Xq11.1 to Xq11.2. This deletion includes a single exon and the 5’ end of each of three isofoms of the gene ARHGEF9, as well as the micro-inhibitory RNA 146B. X-inactivation studies on peripheral blood did not indicate the presence of significant skewed X-inactivation. ARHGEF9 codes for collybistin, a brain-specific GEF which interacts with gephrnin, a key protein of the scaffolding system of inhibitory synapses, that is essential for postsynaptic clustering of both GABA and glycine receptors. Gephyrin has been implicated in genetic risk for autism, schizophrenia and epilepsy. Point mutations, disruptions, and deletions of ARHGEF9 have been reported previously in five patients and are associated with developmental delay, intellectual disability, epilepsy, hyperplesia and mild dysmorphic features. In our patient, the given clinical features are consistent with the deletion of ARHGEF9 associated with haploinsufficiency in the presence of non-skewed X inactivation. Our patient is the first reported case of ARHGEF9 deletion with autism spectrum disorder. This case strengthens the role of ARHGEF9 in cognitive development and its involvement in a subset of patients with neurodevelopmental disorders. The single gene deletion affecting collybistin may lend itself to benefit by treatment with GABA analogues in these patients. In conclusion, ARHGEF9 deletion should be considered in patients with autism spectrum disorder, further investigative studies are warranted to explore the treatment options.
2765F
A four generation family with a novel Hras mutation and predominantly ectodermal findings of Costello syndrome. D. Earl1, S. Wallace1, A. S. 2) Genetic Medicine, Seattle Children’s Hospital, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, WA, USA.
Costello syndrome is characterized by feeding issues and failure to thrive in infancy, short stature, developmental delay, coarse facial features, sparse curly hair, and loose and/or thickened skin. Cardiac findings include congenital heart defects, cardiac hypertrophy, and arrhythmia. Individuals have a 15% lifetime risk for malignant tumors. Hras is the only gene currently known to be associated with Costello syndrome. Germline mutations leading to amino acid substitutions of the glycine residue at positions 12 or 13 are typical. Most probands have a de novo mutation. We present a four generation family with a novel missense Hras mutation and predominantly ectodermal findings. The patient presented at 2.5 years of age with short stature and failure to thrive (-3 SD). Her head circumference was normal (0.3 SD). Birth parameters were normal. Video swallow study showed oral phase dysphagia, microspiration, and gastroesophageal reflex. She had mild obstructive apnea and mild central apnea. Bone age was mildly delayed. Her IGf-1 was borderline low. IGf BP-3 and thyroid function studies were normal. She walked and used her first words at 1 year of age. At 5 years of age, she continues to meet normal developmental milestones. Physical examination revealed proportionate short stature, wooly hair, sparse and kinky eyebrows, piosis with downward eyelashes, epicanthal folds, a wide nasal root, hyperplastic nails, and an enameleastic rash involving cheeks, eyebrows and extremities. On ophthalmologic evaluation she had mild anisometropia and possible left amblyopia. Echocardiogram showed a patent foramen ovale and no evidence of hypertrophic cardiomyopathy at 4 years of age. Abdominal imaging has been normal. Complete function assay was normal despite a report of easy bruising. The patient’s mother, maternal grandfather, maternal great-grandfather, maternal uncle and his son have woolly hair, inflammatory skin findings with hyperkeratosis, nail abnormalities, and normal stature. This indicates that the family history can be explained by incomplete penetrance. Interestingly, the two cousins’ late grandfather also had unilateral renal agenesis and one of the cousins had a brother, who died shortly after a preterm birth having complete renal agenesis. It indicates that the causative mutation could be inherited in a fashion where males also have renal abnormalities. To reveal the etiology of the two cousins’ MRKH syndrome and renal abnormalities, we use a genome-wide approach by carrying out Next-Generation Sequencing of their exomes. We also analyze exomes of the relatives participating in the study. By using a genome-wide approach, we analyse all candidate genes earlier investigated and furthermore have the opportunity of finding new genes associated with MRKH syndrome. We present one of the first studies using a genome-wide approach and the first study using exome sequencing in the investigation of MRKH patients.

2766W
Costello syndrome is characterized by feeding issues and failure to thrive in infancy, short stature, developmental delay, coarse facial features, sparse curly hair, and loose and/or thickened skin. Cardiac findings include congenital heart defects, cardiac hypertrophy, and arrhythmia. Individuals have a 15% lifetime risk for malignant tumors. Hras is the only gene currently known to be associated with Costello syndrome. Germline mutations leading to amino acid substitutions of the glycine residue at positions 12 or 13 are typical. Most probands have a de novo mutation. We present a four generation family with a novel missense Hras mutation and predominantly ectodermal findings. The patient presented at 2.5 years of age with short stature and failure to thrive (-3 SD). Her head circumference was normal (0.3 SD). Birth parameters were normal. Video swallow study showed oral phase dysphagia, microspiration, and gastroesophageal reflux. She had mild obstructive apnea and mild central apnea. Bone age was mildly delayed. Her IGf-1 was borderline low. IGf BP-3 and thyroid function studies were normal. She walked and used her first words at 1 year of age. At 5 years of age, she continues to meet normal developmental milestones. Physical examination revealed proportionate short stature, wooly hair, sparse and kinky eyebrows, piosis with downward eyelashes, epicanthal folds, a wide nasal root, hyperplastic nails, and an enameleastic rash involving cheeks, eyebrows and extremities. On ophthalmologic evaluation she had mild anisometropia and possible left amblyopia. Echocardiogram showed a patent foramen ovale and no evidence of hypertrophic cardiomyopathy at 4 years of age. Abdominal imaging has been normal. Complete function assay was normal despite a report of easy bruising. The patient’s mother, maternal grandfather, maternal great-grandfather, maternal uncle and his son have woolly hair, inflammatory skin findings with hyperkeratosis, nail abnormalities, and normal stature. This indicates that the family history can be explained by incomplete penetrance. Interestingly, the two cousins’ late grandfather also had unilateral renal agenesis and one of the cousins had a brother, who died shortly after a preterm birth having complete renal agenesis. It indicates that the causative mutation could be inherited in a fashion where males also have renal abnormalities. To reveal the etiology of the two cousins’ MRKH syndrome and renal abnormalities, we use a genome-wide approach by carrying out Next-Generation Sequencing of their exomes. We also analyze exomes of the relatives participating in the study. By using a genome-wide approach, we analyse all candidate genes earlier investigated and furthermore have the opportunity of finding new genes associated with MRKH syndrome. We present one of the first studies using a genome-wide approach and the first study using exome sequencing in the investigation of MRKH patients.

2767T
Investigations on the molecular genetic etiology of Mayer-Rokitansky-Kuster-Hauser syndrome in two cousins using exome sequencing. M. Herlin1, S.P. Jonstrup1, A.T. Højland1, I.S. Pedersen2, P.H. Medsen2, A. Ernst3, H. Okkels2, V.O. Le4, H. Kraup3, M.B. Petersen1. 1) Department of Clinical Genetics, Aalborg University Hospital, Aalborg, Denmark; 2) Section of Molecular Diagnostics, Department of Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark; 3) Department of Molecular Genetics and Genetics, Aalborg University Hospital, Aalborg, Denmark.
Mayer-Rokitansky-Kuster-Hauser syndrome (OMIM #277000) is a congenital disorder characterized by aplasia of the uterus and the upper two-thirds of the vagina in females showing normal secondary sex character- istic development, primary amenorrhea, a 46,XX karyotype, and the diagnosis during adolescence when patients present with primary amenorrhea, but they are also troubled with sexual dysfunction and infertility. MRKH syndrome is classified as type 1 (typical/isolated MRKH), type 2 (complex MRKH), type 3 (MRKH with skeletal, or urologic abnormalities), and type 4 (PSSC association). (Müllerian duct aplasia, renal aplasia, and cervical somite dysplasia). The incidence of MRKH syndrome is estimated in 1 in 4,500 live female births. For a long time, the MRKH syndrome was thought to occur sporadically, but cases of familiar clustering led to the hypothesis of an autosomal dominant inheritance with incomplete penetrance. Several candidate genes involved in early embryonic development, such as WNT4, LHx1, WT1, TCF21 (HNF1j) and homeobox genes (HOXA7-HOXA13), have been investigated in earlier studies, but the molecular genetic etiology of the MRKH syndrome remains unknown. We present a familial case of two female cousins (aged 27 and 19) both with type 2 MRKH (with unilateral renal agenesis) that supports the idea of an autosomal dominant heredity. According to basic Mendelian genetics both their mothers, who are siblings, must be carriers of the MRKH causing mutation and some form of penetrance can be explained by incomplete penetrance. Interestingly, the two cousins’ late grandfather also had unilateral renal agenesis and one of the cousins had a brother, who died shortly after a preterm birth having complete renal agenesis. It indicates that the causative mutation could be inherited in a fashion where males also have renal abnormalities. To reveal the etiology of the two cousins’ MRKH syndrome and renal abnormalities, we use a genome-wide approach by carrying out Next-Generation Sequencing of their exomes. We also analyze exomes of the relatives participating in the study. By using a genome-wide approach, we analyse all candidate genes earlier investigated and furthermore have the opportunity of finding new genes associated with MRKH syndrome. We present one of the first studies using a genome-wide approach and the first study using exome sequencing in the investigation of MRKH patients.

2768F
Novel patient with cutis laxa, fat pads and retinopathy due to a novel P5CS mutations and review of the literature. E.V. van Asbeck1, D.F.J. Wolthuis1, M. Mohamed1, T. Gardeitchik1, E. Morava2,1, 1) pediatrics, Radboud University Nijmegen Medical Center, Nijmegen, Netherlands; 2) pediatrics and human genetics, Hayward Genetics Center, Tulean University Medical School, USA.
Autosomal recessive cutis laxa (ARCL) is a connective tissue disorder characterized by wrinkled, inelastic skin; frequently associated with a multisystem disease. We performed next generation sequencing in genetically unexplained patients with neurodevelopmental involvement to assess the underlying etiology. Here, we describe 18 of our patients; diagnosed with a novel, homozygous nonsense mutation in exon 18 of the ALDH1A1 gene, and review all reported P5CS patients. Our patient had feeding issues, progeria-like facial features and severe, generalized sagging skin with visible veins on the abdomen and thorax. In addition, the patient had severe neurological symptoms, including seizures and spasticity. MRI showed microcephaly and hypoplastic corpus callosum. The patient also suffered from pharyngeal dysmotility, delayed fontanel closure, adducted thumbs, salt and pepper retinopathy and failure to thrive. The skin abnormalities were comparable to those described in congenital glycosylation defects (CDG). He had abnormal fat pads on the buttocks and upper legs, and the excessive skin wrinkling, late closing fontanel and seizures are also well known features of CDG. However, ATP6V0A2-CDG and COG7-CDG were ruled out on a molecular and biochemical basis. The adducted thumbs, late closing fontanel, blue sclerae and parchment-like feature of the skin could be due to a PVR1 mutation, although the parchment-like feature of the skin is typically more severe in PVR1 compared to our patient. PVR1 too, was ruled out on molecular basis. However, a mutation in ALDH1A1 was found. So far 7 patients with mutations in ALDH1A1 were described. All patients had failure to thrive, five had short stature, three had inguinal hernia, five had cutis laxa on hands and feet, five had blue veins on the upper leg and seven had CCG and six had joint laxity. Four never learned to walk and four never to speak. Unique features included IUGR, corneal clouding, scoliosis, tremor and dystonia. Our patient is unique due to his abnormal fat distribution, the severe skin findings, the delayed fontanel and severe neurological presentation. Retinitis pigmentosa has never been described in ALDH1A1 patients. In conclusion, the clinical phenotype caused by ALDH1A1 mutations is widely variable. Therefore it is difficult to decide which diagnostic screen for P5CS. We suggest genetic testing for possible ALDH1A1 mutations in patients with cutis laxa, FTT, microcephaly and joint laxity.
2769W
Missense mutation of MAF in a Japanese family with congenital cata-
ract. Y. Narumi1,2, S. Nishina3, M. Tokimitsu2, Y. Aoki2, R. Kosako2, T.
Kosho3, T. Murata2, F. Takada2, Y. Fukushima2,1 1) Department of Medical
Genetics, Kitasato University Graduate School of Medical Sciences, Sagam-
ihara, Kanagawa, Japan; 2) Division of Ophthalmology, National Center for
Child Health and Development, Tokyo, Japan; 3) Department of Ophthalmol-
ogy, Shinshu University School of Medicine, Matsumoto, Japan. 1) Depart-
ment of Medical Genetics, Tohoku University School of Medicine, Sendai,
Japan; 5) Division of Medical Genetics, National Center for Child Health and
Development, Tokyo, Japan; 6) Department of Medical Genetics, Kitasato
University Graduate School of Medical Sciences, Sagamihara, Japan.

Congenital cataract is defined as a crystalline severe blinding disease and
onset within the first year of life. One-fourth patients with bilateral cataracts
have inherited trait. More than 40 causative genes of cataracts are encoding
structural, cytoskeletal, gap junction channel, membrane associated and
cell signaling proteins. It considered difficult to establish phenotype-genotype
correlations due to genetic heterogeneity. Therefore, comprehensive muta-
tion analysis using whole exome sequencing (WES) is suitable for the diag-
nosis of inherited cataracts. We report a family with congenital cataract and
MAF mutation which were identified by WES. All affected family showed
congenital cataract and some members had with coloboma and microcoro-
nea. Male family members revealed autism spectrum disorder and hernia
inguinal. It was initially considered that the family had an X-linked sex-
dominant inheritance pattern. WES was conducted for 3 DNA samples and
MAF was included of segregated variants. The MAF mutation in the family
was not reported previously. The mutations have not been identified in 200
alleles in Japanese control and other unaffected members. These data
described that the mutation in MAF is likely the causal mutation in this family.
MAF mutation had been reported of only 5 families. The family members
which were previously reported had shown various type of cataract and
rapidly of onset. The patients with MAF mutation shared not only cataracts
but also anterior segment abnormalities. Our recent data strongly support
that congenital/juvenile cataracts with MAF mutations showed microcornea
and coloboma.

2770T
Type II Collagenopathy Patients in Korea. E. Ra1, H. Park2, S.H. Seo3, S.I. Cho3, T.J. Cho4, N. Park1, M.W. Seong1, S.S. Park2. 1) Department of Laboratory Medicine, College of Medicine, Seoul National University Hospital, Seoul, Korea; 2) Division of Pediatric Orthopaedics, Seoul National University Children’s Hospital, Seoul, Korea.

Introduction: Type II collagen is a helical molecule which is found in carti-
lage and the vitreous humor of the eye. Due to its widespread distribution, type II
collagenopathies show a spectrum of significantly variable conditions. Here we present the phenotypic spectrum of Korean type II collagenopathy
patients whose diagnosis was confirmed by the mutation analysis of COL2A1
genome. Methods: Twenty Korean patients who were suspected of type II
collagenopathy were screened for the COL2A1 gene. DNA was extracted
from peripheral blood leukocytes, and COL2A1 gene mutations were
screened via direct sequencing of all 54 exons and multiplex ligation-depen-
dent probe amplification (MLPA). Results: Heterozygous COL2A1 mutation
was identified in 9 cases, which was the most common finding in this study. In 2 cases,
diagnoses were not clearly categorized. Interestingly, a nonsense mutation
which had been previously reported in a case of Stickler syndrome was
identified in one patient with Legg-Calve-Perthes disease, another with family history of early onset osteoarthritis and three with suggestive findings of spondyloepiphysial dysplasia and brachydactyly.

Conclusion: Variable phenotype may occur even within the individuals shar-
ing a similar COL2A1 mutation. The COL2A1 mutation, therefore, is not a
sole causative factor for the type II collagenopathy. In addition, there may be
other genetic factors related in the developmental process. Confirming the
diagnosis of a type II collagenopathy by mutation analysis is important in
predicting the possible complications of the condition as well as providing
an accurate genetic counseling for other family members.

2771F
Clinical spectrum of eye malformations in 4 new patients with Mowat
Wilson syndrome. A. BOURCHANY1, I. GIURGEA1, J. THEVENON1, A. GOLDENBERG2, G. MORIN2, D. BREMOND-GIGNAC2, P-O. LAFON-
TAINNE1, D. THOUVENIN2, J. MASSY2, A. MASUREL-PAULET2, C. THAU-
VIN-ROBINET3, S. EL CHEHADA3, S. LYONNET2, L. FAIVRE2. 1) CHU
Hôpital d'Enfants, University of Burgundy, Dijon, France; 2) Department of Genetics, Hôpital Henri Mondor, University of Paris Est Creteil, Paris, France; 3) Department of Genetics and Rare Diseases and Development Abnormalities and Malformative Syndromes, CHU Hôpital d'En-
fants, University of Burgundy, Dijon, France; 4) Department of Clinical
Genetics, Hôpital Charles Nicolle, University of Rouen, Rouen, France; 5) Department of Oncogenetics and Clinical Genetics, hôpital Nord, University of
Picardie Jules Verne , Amiens, France; 6) Pediatric Ophthalmology Saint Victor Center, University of Picardie Jules Verne, Amiens, France; 7) Depart-
ment of Medical Genetics, Hôpital Necker-Enfants Maladies, University
of Paris 5 René-Descartes, Paris, France; 8) Jean Jaurès Ophthalmic Center,
Toulouse, France; 9) Hôpital Charles Nicolle, University of Rouen,
Rouen, France.

Mowat-Wilson syndrome (MWS) is a multiple congenital anomaly syn-

drome characterized by a distinct facial phenotype, intellectual deficiency,
Hirschprung disease and multiple malformations caused by mutations or
deletions in the zinc finger E-box-binding homeobox2 gene (ZEB2). To date,
neither a congenital lens anomaly was reported in the literature. The ZEB2
coding a transcriptional repressor involved in the transforming growth factor β (TGFβ) signaling pathway and is strongly tran-
scribed at an early stage in the developing central and peripheral nervous
system in mice. Eye abnormalities have been rarely described in patients
with this syndrome (4% in literature). Herein, we describe 4 patients (2
males and 2 females) with MWS and eye malformations, with a mean age of
7 years. Eye malformations included microphthalmia, iris/retinal colobomas,
atrophy or absence of optic nerve, absence of optic disc, total hyphema,
and deep refraction troubles, sometimes with severe visual consequences.

Malformations were asymmetric in all cases and often unilateral. This spec-
trum of manifestations, affecting all eye segments, was inconstant with
the previous cases mentioned in the literature (iris/chorioretinal/optic disc
cryptogenesis, cataract, nerve atrophy, optic nerve atrophy, microphthalmia,
korectopia), as well as expression data of ZEB2 in human embryo. With the help
of the literature data, we questioned whether or not the presence of eye
manifestations could be due to specific type or locations of mutations.
We showed that the presence of eye malformations, although a rare feature in
MWS, should not rule out the diagnosis.
2773T
Weaver Syndrome: Variable expression and natural history in a three generation family with documented EZH2 mutation. H. Hoyne1, A.R. Mroch1, M. Guyot2, L. Hatakeyawa2, S. Lee2, L.H. Seaver1
1) Sanford Research/Office of Academic Affairs, Sanford Health, Sanford School of Medicine of the University of South Dakota, Sioux Falls, SD; 2) Kapioiian Medical Specialists, University of Hawaii John A. Burns School of Medicine, Honolulu, HI
Weaver syndrome is an autosomal dominant condition characterized by tall stature, macrocephaly, round face in infancy, broad forehead, hypertelorism, large ears and early developmental delay. Recently, mutations in the EZH2 gene have been causative for the phenotype. This report is to provide detailed information on the natural history and intragenic phenotypic variability in six individuals in a 3 generation family with Weaver syndrome, including the oldest reported patient, a 53 year old female. Our proband, a 10 year old girl, was evaluated for overgrowth at 10 months with a family history of Sotos syndrome. By age 2.5 years, her features were noted to be more typical of Weaver syndrome. Her mother and sister were evaluated at ages 26 and 2 years respectively, for overgrowth. The mother is 188 cm tall with an OFC of 59 cm. As an adult, the proband’s sister is 183 cm tall and has 4 children, with overgrowth and features suggestive of either Weaver or Sotos syndrome. The two eldest had normal NSD1 testing. EZH2 testing on the two youngest children (both clinically affected) revealed a previously identified pathogenic mutation, c.2050C>T, p.R684C. The mutation was confirmed in our proband. Mutation analysis on the two older children, one clinically affected and one clinically normal, is pending. Term birth weights ranged 3.1-5.2 kg. All had excessive linear growth and macrocephaly during infancy, speech delay and articulation deficits. At school, they have motor delays. The results indicate severe cognitive delay in school. Medical complications include scoliosis, duplicated ureter, avascular necrosis requiring early hip replacement, pulmonary embolism, tansilestis for airway obstruction, cholecystectomy, umbilical hernia repair and neonatal tachyarrhythmia. While the symptoms were congenital, they have had malignancy. Several of the children demonstrated a phenotype more compatible with Sotos syndrome in infancy, progressing to a Weaver syndrome phenotype as they aged. Data from this family further enlarge the phenotype of the Weaver syndrome and highlight the complexity of distinguishing Weaver from Sotos syndrome in isolated cases. The identification of EZH2 mutation as causative for Weaver syndrome will help differentiate the two disorders and allow for more accurate phenotypic characterization of both.

2774F
Xeroderma pigmentosum complementation group B / Trichothiodystrophy spectrum in two siblings with ERCC3 mutations. M. Miglia-vacca1, N. Sobreira2, S. Bragagnolo1, M. Ramos1, D. Vallo2, A. Perez1
1) Clinical Genetics, Federal University of São Paulo, São Paulo, São Paulo, Brazil; 2) Institute of Genetic Medicine , Johns Hopkins University School of Medicine, Baltimore, MD, USA
Nucleotide excision repair (NER) is the main cellular repair pathway by which a wide range of DNA lesions are eliminated from the genome. The importance of DNA repair is demonstrated by the existence of at least 15 human genetic disorders associated with defects in DNA repair and defects in appropriate response to DNA damage. NER is involved in cancer incidence and multi-system defects. Xeroderma pigmentosum (XP, OMIM 278700-278780), Cockayne syndrome (CS, OMIM 216400, 133540) and Trichothiodystrophy (TTD, OMIM 601675) are autosomal recessive disorders caused by defects in the repair of DNA, remarkable clinical heterogeneity exists within these diseases. The ERCC3 gene (Excision-Repair Cross-Complementing group 3) is related to them and encodes a DNA helicase that acts in the NER pathway and a component of the transcription factor BTF2/TFIIH that initiates transcription of class II genes. A missense mutation in the NER pathway can result in the classic phenotype of XP, while inactiva-
tion of transcription can lead to photosensitivity phenotype like XP, CS and TTD. Here we describe a sister and a brother born to a non consanguineous 26 years old mother and a 22 years old father, both with developmental delay, recurrent pneumonia, cutaneous xerosis, moderate sensorineural hearing loss, and retinitis pigmentosa. The sister had head circumference of 53 cm (<p10), her weight was 4.7 kg (p3) and height was 1.56 cm (p10), she had dry hair, sparse eyebrows, and a skin lesion in the right anterior cervical region with a descamative aspect. The brother’s head circunference was 53 cm (<p3), his weight was 67 kg (p25-50) and height was 1.67 cm (p10), he had microcephaly, dry hair and ichthyosis. To date no skin tumors were diagnosed in neither one of them. WES was performed in the brother and sister. Sequencing showed that the mother was compound heterozygous for both mutations. Our results imply that the ERCC3 gene as the responsible for the phenotype in this family. Up to now only six families have been described in the literature with mutations in ERCC3. The absence of skin cancer in our patients may be explained by compensation through an alternative process and by the nature of the compound heterozygous mutation.

2775W
Craniofacial phenotypes in cutis laxa. Z. Urban1, C. Lorenchick1, T.E. Parsons2, K. Levine3, S. Madan-Khetarpal2, S.M. Weinberg1
1) Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 2) Department of Oral Biology, University of Pitts-
sburgh School of Dental Medicine, Pittsburgh, PA; 3) Department of Pediatric, University of Pittsburgh School of Medicine, Pittsburgh, PA
Cutis laxa (CL) is a rare connective tissue disorder. Anecdotal reports suggest that individuals with CL may have distinctive craniofacial features, but no quantitative analysis of the craniofacial phenotype has been carried out on this population. Our goal was to apply objective methods to this problem. 23 subjects with cutis laxa were included in this study. Participants underwent physical examination and the medical history and results of the examination were collected using a questionnaire. 3-dimensional (3D) images of the head and face were acquired using a 3DMD portable stereo-photogrammetry system. 24 facial landmarks were identified on each sub-
ject’s 3D facial image, and a geometric morphometric analysis was per-
formed to quantify facial shape. A subset of 6 CL subjects were compared to 18 sex and age-matched controls by principal components analysis. The first principal component captured 39% of the total variation and separated CL subjects from controls. CL patients had significant narrowing and elonga-
tion of the face as well as mandibular hypoplasia. Our studies define and quantify craniofacial traits in CL and implicate CL genes as regulators of craniofacial development.

2776T
Expanding the RAD21 mutational spectrum: report of the first intragenic deletion and frameshift mutation in two patients with a mild form of Cornelia de Lange Syndrome. A. Minor1, M. Shinawari1, J.S. Hogue2, D.R. Hamlin3, C. Tan1, K. Donato1, L. Wyanger1, S. Botes1, S. Das1, D. del Gaudio1
1) Human Genetics, University Chicago, Chicago, IL; 2) Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, Missouri; 3) Depart-
ment of Pediatrics, San Antonio Military Medical Center, Fort Sam Hous-
ton, Texas
Cornelia de Lange syndrome (CdLS) is a developmental disorder charac-
terized by limb reduction defects, characteristic facial features and impaired cognitive development. In the majority of cases mutations in NIPBL can be identified. However, additional genes have been implicated in the disorder. Recently, missense mutations and whole gene deletions in RAD21 have been identified in children with growth retardation, minor skeletal anomalies, and facial features that overlap findings in individuals with CdLS. We expand the RAD21 mutational spectrum by reporting the first intragenic deletion and the first frameshift mutation identified in two patients presenting with a mild form of CdLS. Patient 1 is a 2.5 year old boy with developmental delay, hypoplasias, inguinal hernia and dysmorphic features that included mild synphysis, coarse facial features, 5th finger clinodactyly and overlapping of toes two over three. Patient 2 is a 12 year old boy who was diagnosed with CdLS at around a year of life on the basis of developmental delays, characteristic facial features, hirsutism, and hand and feet anom-
alia. In patient 1, an intragenic deletion involving RAD21 exon 13 was identified by exon-targeted array comparative genomic hybridization. Breakpoint junction sequence analysis identified a five base pair region of microhomol-
gy around the deletion breakpoints, suggesting that the deletion arose through a microhomology-mediated repair mechanism. In patient 2, sequence analysis of the RAD21 coding region revealed a two base-pair duplication in exon 6, c.592_593dup. This frameshift duplication causes a premature stop codon at amino acid position 203. mRNA sequences that contain a premature stop codon may be targeted for nonsense mediated decay. The single exonic deletion and the frame shift mutation of RAD21 identified in our patients were associated with a mild presentation of CdLS. As supported by previous literature reports, our data confirm that RAD21 mutation analysis is indicated in patients presenting with milder phenotypes of CdLS.
Two Novel Mutations in a Patient with Rhizomelic Chondrodysplasia Punctate Type 1. H. Onay1, M. Saka Guvenç1, T. Atik2, A. Aykut1, O. Cogulu3,4, F. Ozturk5,6, 1) Medical Genetics, Ege University Faculty of Medicine, Izmir, Turkey; 2) Pediatrics, Ege University Faculty of Medicine, Izmir, Turkey.

Rhizomelic chondrodysplasia punctate (RCDP) is an autosomal recessive peroxisome biogenesis disorder characterized by prominent skeletal manifestations, such as proximal shortening of the long bones, multiple punctate epiphyseal calcifications and multiple joint contractures. Additionally, these patients have congenital cataracts, severe mental deficiency and postnatal growth retardation. Dismorphic features such as hypertelorism, midface hypoplasia, small nose and full cheeks are seen in these patients. Three clinically indistinguishable forms of RCDP have been defined up to now and mutations in three different genes are responsible for these specific forms. PEX7 gene is responsible for the most common form of the disease, Type 1 RCDP. Only a few patients survive beyond first decade and death usually results from generalized neurologic impairment and respiratory complications. PEX7 gene, which comprises 10 exons is located on 6q21-q22. Due to mutations of this gene, transportation of PTS2 containing proteins including peroxisomal 3-ketoacyl-CoA thiolase and PTS2-EFGP protein to the surface of the peroxisome is impairedd. Herein we present a case with dismorphic features such as depressed nasal bridge, short philtrum, high arched palate, antverted nares; shortened femur,ibia,humerus and fibula which were diagnosed radiologically, multiple punctate epiphyseal calcifications and congenital cataract. There is consanguinity between parents. Mutation screening in the PEX7 gene revealed two new mutations that were never identified before. p.D308Y mutation which is located in exon 10 was inherited from mother and IVS5-2A>G was inherited from father. In silico analysis indicated that these two mutations were harmful. We present this case to demonstrate two novel mutations detected in PEX7 gene and to remind the importance of having two different mutations in consanguineous families.

Hereditary multiple osteochondromas. Molecular characterization of three Cypriot families and report of two novel EXT1 gene deletions. G.A. Tanteles1, V. Nioleou2, C. Shammas3, E. Ellina4, L.A. Phylactour1, C. Sismani5, V. Anastasiadou-Christophidou1. 1) Clinical Genetics, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; Cyprus; 2) Department of Molecular Genetics, Function and Therapy, The Cyprus Institute of Neurology & Genetics, 1683, Nicosia, Cyprus; 3) Department of Cytogenetics and Genomics, The Cyprus Institute of Neurology & Genetics, 1683, Nicosia, Cyprus.

The disorder hereditary multiple osteochondromas (HMO) - previously called hereditary multiple exostoses (HME) - is characterized by the development of multiple osteochondromas (benign cartilaginous bony tumors which grow outward from the metaphyses of long bones) which can lead to a reduction in skeletal growth, bony deformities, restricted joint motion, short stature, premature osteoarthrosis, and compression of peripheral nerves. The risk for malignant transformation (typically osteochondrosarcoma) increases with age, although the lifetime risk is low (~1%). HMO is inherited in an autosomal dominant manner and is caused by mutations in either the EXT1 or EXT2 gene. Malignant transformation has been described in about 70% of cases and in 20-30% of cases, respectively. This study presents the molecular characterization of three Cypriot HMO patients and their families who underwent genetic analysis of the EXT1 and the EXT2 genes. Mutation analysis was performed by direct sequencing of the entire coding regions of both genes. In an attempt to identify deletions or duplications, MLPA and array-CGH analyses were also used. Two novel EXT1 gene deletions and the previously known EXT1 p.R701stop mutation were respectively identified in three probands and members of their families. The two novel EXT1 mutations were detected in a 4.64kb to 0.16Mb. The breakpoints of the first deletion were located within the coding region of exon 1 of the EXT1 gene while the second deletion removed exons 2 to 11 of the same gene. These observations provide additional evidence of the variability in phenotypic expression and mutational spectrum which is characteristic of HMO, and present the first genetic analysis of Cypriot HMO patients. Hopefully this analysis can help towards effective diagnosis, improved genetic counseling and be used as a future potential therapeutic platform for patients affected with HMO in Cyprus.

First molecular study of Kindler syndrome in three Iranian families: novel and recurrent mutations in the FERMT1 gene. H. Vahidnejad1,2, L. Youssefi3, A. Yaqunfar1, A.M. Kajbafzadeh4, F. Agha-Hosseini2, H. Tabrizi2. 1) Biotechnology Research Center, Department of Molecular Medicine, Pasteur Institute of Iran, Tehran, Iran; 2) Medical Genetics, Tehran University of Medical Sciences, Tehran, Iran; 3) Department of Dermatology, Hamedan University of Medical Sciences, Hamedan, Iran; 4) Pediatric Urology Research Center, Pediatric Center of Excellence, Tehran University of Medical Sciences, Tehran, Iran; 5) Department of Oral Medicine, Dental Research Center, Dentistry School, Tehran University of Medical Sciences, Tehran, Iran.

Bullous acrokeratotic poikiloderma or Kindler syndrome (OMIM 173650) is a rare autosomal recessive genodermatosis characterized by dysphagia, extensive diffuse poikiloderma, minimal trauma-induced blisters, acrokeratosis with cutaneous atrophy and cutaneous atrophy. Novel mutations in the FERMT1 gene encoding kindlin-1, a component of focal adhesions in keratinocytes in which less than 60 mutations have been reported so far. In contrast to most other countries, the burden of this disease seems to be higher in Iran due to consanguineous marriages. For the first time, we conducted a molecular analysis of three separate families from different ethnic populations in Iran toward better genetic characterization and better management of this unique population. After obtaining written informed consent from families, extensive family history was documented and DNA was extracted from patient blood samples. PCR reaction was set up for 15 exons and nearby intronic regions of the FERMT1 gene. PCR products from eight patients in three separate families were bidirectionally sequenced and analyzed. Two affected brothers from central Iran harbored the c.328C>T (p.Arg110X) homozygous mutation in exon 3. Extensive mutation analysis was also performed for four other siblings in this family, and in genetic counseling sessions the available reproducible options and recurrent risks for the heterozygotes were discussed. From Azerbaijan, two affected brothers presented with the least severe clinical manifestations with only dysphagia and no extra-cutaneous involvement. The two Azerbaijani brothers seem to be homozygotes for a probable missense mutation c.1577G>A (p.Arg526Lys) after sequencing and analysis of the entire length of the gene. This variant, in the clinical form, is classified as a polymorphism and non-pathogenic. This is the first report of homogygosity and pathogenicity of this variant. A third family from western Iran with four severely affected siblings were found to harbor a novel mutation c.1383C>A (p.Tyr461X). In conclusion, we provide evidence for novel mutations in the FERMT1 gene from Iran and expand the existing spectrum of mutations in humans. This data also allows us to provide genetic testing to affected individuals and propose the possibility of using PCR and sequencing as a cheap and quick screen of the heterozygotes in the population. Genetic counseling, education, and follow-up could reduce the burden of this disease in Iran.

Genotype-Phenotype Relationships in Freeman-Sheldon Syndrome. L. Youssefi1,2, M. Dekkers3, P. Nies, H. Kibade, M. Rona, M. M. Greeson, M. stylish, S. H. Gaffney, H. Bamshad1,2,3. 1) Department of Pediatrics, Division of Genetic Medicine, University of Washington School of Medicine, Seattle, WA; 2) Seattle Children’s Hospital, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA.

The distal arthrogryposis (DA) syndromes are a group of ~10 disorders characterized by multiple congenital contractures such as camptodactyly and clubfoot. DA type 2A (i.e., Freeman-Sheldon syndrome) is the most severe of the DA syndromes; as it is also marked by contractures of the facial muscles. DA2A is caused by autosomal-dominant mutations in MYH3, the gene that encodes embryonic myosin heavy chain, a protein of the contractile apparatus of skeletal muscle. In our cohort of 46 families with DA2A, we found disease-causing mutations in MYH3 in 43 kindreds (93%). We then characterized the genotype-phenotype relationships in these mutation-positive families with DA2A. Three common MYH3 mutations accounted for 39/43 (91%) of the DA2A mutation-positive kindreds including: 8/43 (19%) with p.R672H, 11/43 (26%) with p.R672C, and 20/43 (47%) with p.R672H. Using a quantitative phenotype score that we developed based on physical findings and natural history, we found individuals with p.T178I to have a phenotype score that was more severe than individuals with a p.R672C or p.R672H mutation. Specifically, the p.T178I mutation was associated with more severe facial contractures and congenital scoliosis. The p.R672C mutation was associated with less severe foot and facial contracture scores. Of the total 46 cases screened, 11 were familial cases and 35 were sporadic cases. MYH3 mutations were found in all of the familial cases but only 8/11 (73%) familial cases. Specific mutations were more common among sporadic cases than familial cases. For example, the mutation with the most severe phenotype scores, p.T178I, was observed in only sporadic cases (8/8). The mutation associated with the least severe phenotype scores, p.R672C, was observed only in 11/35 (27%) sporadic cases but 8/11 (73%) familial cases. Collectively, our analysis suggests that MYH3 genotype is predictive of phenotype and, to a more limited extent, natural history of individuals with DA2A.
2781W
A 24bp deletion in ELN causing a Marfan-like phenotype. J. Hoyer, C. Kraus, A. Reis. Institute of Human Genetics, University Erlangen-Nuremberg, Erlangen, Germany.

We report on a 27 year old male presenting with a Marfan-like phenotype. The patient had a mild myopia, a high palate and broad uvula, skin striae, a flat foot, a positive wrist sign as well as long fingers and toes and an increased arm span-to-height ratio. Moreover a threefold patellar laxation was reported. Marfan syndrome was suspected three years ago when funnel chest surgery was performed. A cardiological examination was consecutively recommended but revealed no abnormalities. The recently revised diagnostic criteria for Marfan syndrome integrate information from multiple sources including personal medical history and physical examination. Those features are weighted and grouped to derive a ‘systemic score’. In the absence of a family history of Marfan syndrome, a systemic score of 27 in combination with an aortic root enlargement is sufficient for the diagnosis. Our patient achieved a systemic score of 7 but did not show an aortic dilatation. We used an individual, Multiplex-PCR based Ion AmpliSeq Kit capturing 98.44 per cent of coding regions to analyze the genes FBNA, ACTA2, ELN, CBS, FBG2, MYH11, COL3A1, SLCA10, SMAD3, TGFBR1, TGFBR2 and TGFαB. Surprisingly we identified a small heterozygous in frame deletion of 24 bp (c.1178_1201del24bp) in exon 20 of the ELN gene leading to a loss of 8 amino acids (p.Gly393_Ala401delinsAla). This result was confirmed by Sanger sequencing. We suppose that expression of a mutated protein leads to the disruption of elastic fibre architecture. Up to now mutations in Elastin are only known for autosomal dominant Cutis laxa (ADCL) and as a cause for aortic aneurysms but are not described as causative for other Marfan phenotype features. For further evaluation physical examination and molecular analysis in the patient’s parents is planned.

2782T
Multiple pterygium syndrome, Escobar variant, in a patient with a congenital diaphragmatic hernia and prenatally-diagnosed arthrogryposis. A.L. Sutton, M.D. Descartes. Department of Genetics, University of Alabama at Birmingham, Birmingham, AL.

Introduction: The multiple pterygium syndromes (MPS) are characterized by arthrogryposis, webbing of the skin (pterygia), dysmorphic facial features, and other congenital anomalies. There are two forms of MPS: the lethal and non-lethal (Escobar) variant. MPS displays significant locus heterogeneity. Some cases are caused by mutations in the gene encoding the gamma-2 chain of type III collagen (COL3A1). Others have a congenital diaphragmatic hernia with MPS, Escobar variant, with heterozygous mutations in CHRNG. The mother, a 40 year old G3P2, was referred to our fetal diagnosis center at 25 weeks of gestation for congenital eventrations have been found in patients with MPS, Escobar variant, with heterozygous mutations in CHRNG. The mother, a 40 year old G3P2, was referred to our fetal diagnosis center at 25 weeks of gestation, with multiple congenital anomalies. The detailed ultrasound showed multiple joint contractures, skin edema, and polyhydramnios. Aminoacetonitrite remained a normal female karyotype. She was delivered via cesarean section after a failed induction of labor at 39 weeks. After birth, she was found to have a congenital diaphragmatic hernia, which was repaired on day of life 12. Genetic testing revealed two truncating mutations in CHRNG (c.459dupA and c.753_754delCT). After serial casting and physical therapy, she started standing at 16 months of age. Discussion: Diaphragmatic hernias are one of the findings in lethal multiple pterygium syndrome. Although diaphragmatic hernias have not previously been described in this syndrome. Additionally, homozygosity for these two truncating mutations has been identified with the lethal form of the disorder, while compound heterozygotes for these two mutations with the Escobar phenotype similar to this case have also been described. This case highlights the phenotypic and genetic heterogeneity in the MPS. MPS should be considered in the differential diagnosis in patients with prenatally-diagnosed arthrogryposis and congenital diaphragmatic hernias.

2783F
Congenital Dyserythropoietic anemia type 1 (CDA 1) presenting with blueberry muffin rash and profound anemia. L. Turner1, L. Goodyear2, L. Bowes2, S. Fernandez1, A. Hogg1. 1) Medical Genetics, Memorial University of Newfoundland, St. John’s, NL, Canada; 2) Pediatric Hematology/Oncology, Eastern Health, Janeway Child/Rehab Centre, St. John’s, NL, Canada.

The term blueberry muffin rash was first used to describe a non-blanching purpuric rash that can be seen in neonates with congenital infections. Biopsy findings of the lesions seem reveal extramedullary hemopoesis. The differential diagnosis of blueberry muffin rash includes infection, neoplasms, and hematologic dyscrasias. More recently, Alcicardi-Goutieries syndrome and Mucopoliodosis 2 have been added to the list of disorders that can present in the neonatal period with blueberry muffin rash. We present a 37 week gestation singleton male born with a blueberry muffin rash and hepatomegaly. He was depressed at birth requiring intubation and ventilation. Initial CBC revealed a hemoglobin of 56 g/L. Physical examination revealed several digit abnormalities including small finger and toenails, 4-5 toe soft tissue syndactyly and a small 3rd toe on the right foot. The infant was also noted to have a short neck with redundant nuchal skin, bilateral epicanthic folds and mild micrognathia. He had significant hepatosplenomegaly. Figure inves- tigations for congenital infection were negative. Karyotype and microarray were normal. Ongoing investigations including repeat blood smear and bone marrow examination showed abnormalities but were not diagnostic of any one condition. Given the hematologic picture and the digit abnormalities CDA 1 was considered. Molecular analysis revealed compound heterozygous mutations in the CDAN1 gene. Our patient was started on interferon and is doing well. He is the first reported patient with CDA 1 presenting with blueberry muffin rash. CDA 1 should be added to the differential diagnosis of infants presenting with blueberry muffin rash.

2784W
Whole exome sequencing uncovers mutations in MYH9 associated with expanded phenotype spectrum. C. Rao1, J. Liang2, A. Alodabi2, Y. Guo1, T. Tian1, X. Liu3, L. Dai1, B. Keating4,4, M. Menezes3, W. Gold2, M. Wilson2, L. Ades5, J. Zhang6, A. Kakakios3, J. Wang2,6,9,10, H. Hakonarson1,4, J. Teo1. 1) Center for Applied Genomics, Children’s Hospital of Philadelphia; Philadelphia, PA, USA; 2) BGI-Shenzhen; Shenzhen 518083, China; 3) Genetic Metabolic Disorders Research Unit, Children’s Hospital at Westmead; Westmead, New South Wales, Australia; 4) Division of Human Genetics, The Children’s Hospital of Philadelphia; Philadelphia, PA, USA; 5) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania; Philadelphia, PA, USA; 6) Clinical Genetics Dept, Children’s Hospital at Westmead; Westmead, New South Wales, Australia; 7) Immunol- ogy Dept, Children’s Hospital at Westmead; Westmead, New South Wales, Australia; 8) Department of Biology, University of Copenhagen; Copen- hagen, Denmark; 9) King Abdulaziz University; Jeddah, Saudi Arabia; 10) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark; Copenhagen, Denmark; 11) Haematology and Blood Cancer, Children’s Hospital at Westmead; Westmead, New South Wales, Aus- tralia.

Dominantly inherited deleterious mutations in the human gene MYH9 (myosin, heavy chain 9, non-muscle) have been linked to a range of rare hereditary disorders with the overlapping platelet macro- thrombocytopenia (MTPC), and they can also cause associated symptoms of hearing impairment, nephropathy, leukocyte inclusion and/or cataract. In the current study, we used whole exome capture followed by next generation sequencing to search for causative mutations in two families with MTPC. The proband in the first family presented with MTPC, craniostenosis and hearing loss and was found to have a MYH9 mutation (amino acid change S96L); the proband in the other family presented with MTPC, severe mental retardation, epilepsy, facial deformations and renal calculi and was found to have another MYH9 mutation (A9ST). While these two mutations have been reported before, the presentation of developmental, intellectual and psychiatric abnormalities is atypical and new to the MYH9-related diseases. We validated these mutations by traditional Sanger sequencing and the WES data excluded the possibility of other genetic mutations being causal for the syndromes. These results demonstrate that by providing fast and accurate diagnosis at the molecular level, next generation exome sequencing can help broaden the symptomatic spectrum of an established genetic disorder when the presentation is atypical.
V. Cormier-Daire, E. Salort-Campana

caused by autosomal dominant mutations in FAM111B
ty and pulmonary fibrosis, a phenotypically recognisable syndrome, is
syndrome at first sight. In conclusion, HFP with tendon contracture, myopa-

negative mutations resulting in FAM111B enzymatic activity changes. Func-
tional studies are ongoing to better understand the pathophysiology of this
gene (NM_198947.3), unreported to date. The involvement
African family and in the trio of one sporadic case gives rise to the identifica-
MRI shows extensive fatty infiltration confirmed by muscle biopsy. Micros-

6-30). Key features consist of: (i) congenital poikiloderma, hypotrichosis,

sive muscular weakness and (iv) progressive pulmonary fibrosis. Muscle

tional and pulmonary infections. Mutations in STAT3 cause autoso-
dominant (AD) HIES and mutations in the DOCK8 gene are described
in autosomal recessive (AR) HIES. In many HIES cases, the genetic cause
remains unknown. Here we describe a novel immune dysregulation pheno-
type with HIES-like features due to underlying mutations of the MAL1 gene,
a key component of NF-κB signaling, highlighting a new role for the NF-κB
pathway in human disease. Our patient, a 14 year-old girl born to consanguini-
neous parents, presented in infancy with widespread, treatment-resistant
dermatitis, followed by recurrent bacterial and viral skin infections, episodes
of pneumonia, chronic lung disease, severe gastrointestinal inflammatory
pathway in human disease. Our patient, a 14 year-old girl born to consanguini-
neous parents, presented in infancy with widespread, treatment-resistant
dermatitis, followed by recurrent bacterial and viral skin infections, episodes
of pneumonia, chronic lung disease, severe gastrointestinal inflammatory
disease, osteoporosis with fractures, chronic granulation tissue of larynx
and ears, and severe periodontal disease, along with short stature and non-
dynamic facial features. Laboratory testing showed variable eosinophilia, very
B cell numbers, chronically elevated serum IgE and failure of lymphocyte proliferation upon phytohemagglutinin (PHA) stimu-
lation. By whole exome sequencing, we identified an underlying homozygous
missense mutation in the MAL1 gene encoding MAL1/paracaspase, a (modifying NF-κB-related substrates), and key scaffold protein of the
CARMA1-BCL10-MAL1 (CBM) signalling complex (critical for CBM-
mediated NF-κB signalling). We demonstrate that the MAL11 mutation results
in profound deficiency of the NF-κB pathway, associated with silencing of
mediated activation of NF-κB and loss of paracaspase proteolytic activity.
Characterization of human MAL1 deficiency expands our understanding
of the genetic causes of disorders with HIES features and the role the NF-
κB pathway in human immune disease. In addition, as MAL1-inhibitors are
currently being examined for the treatment of lymphoma and inflammatory
conditions, our case provides unique insights to guide the ongoing develop-
ment of these therapies.
2787W
Familial Occurrence of Multiple Isolated Epidermal Inclusion Cysts: evidence for X-linked inheritance? J. Jenkins, K. Horii, H. Ardinger. Divisions of Genetics and Dermatology, Children's Mercy Hospitals & Clinics, University of Missouri-Kansas City School of Medicine, Kansas City, MO.

Epidermal inclusion cysts (EICs) are among the most common type of subcutaneous cysts and are frequently found on the face, trunk, neck and scalp. EICs originate from the infundibular follicle and are characterized by implantation of epidermal elements into the dermis. Most EICs occur as sporadic, isolated lesions and are typically benign with rare malignant transformation reported. EICs are twice as common in males, show no racial predilection, and commonly arise in adulthood. Multiple EICs have been reported in association with underlying genetic syndromes including familial adenomatous polyposis (FAP), pachyonychia congenita, and nevoid basal cell carcinoma syndrome (NBCCS). Autosomal dominant inheritance of multiple isolated EICs has also been suggested (OMIM #131600); although a specific genetic etiology remains unknown. We report a 17 year-old male with a history of multiple subcutaneous cysts involving the trunk, face, neck and buttocks beginning at age 5. The diagnosis of EICs was confirmed on pathology after surgical excision. The family history was notable for early onset of multiple EICs in the patient's mother, maternal aunts, maternal grandmother, maternal great-grandmother, and maternal-great-grandfather. There was no family history of polyposis, cancers, nail dystrophy, cleft craniofacial dysmorphism or other findings suggestive of FAP, pachyonychia congenita, or NBCCS. The family history could be compatible with either an autosomal dominant or an X-linked inheritance pattern of multiple isolated EICs. All daughters of affected males in this pedigree are affected, males are affected more severely than females, and lack of male-to-male transmission are all factors which could support an X-linked inheritance pattern. The recurrence risk for multiple EICs in the patient's offspring ranges from 50% for either gender in the case of autosomal dominant inheritance, up to 100% in female offspring in the case of X-linked dominant inheritance. While autosomal dominant inheritance has been previously suggested in familial cases of multiple isolated EICs, few families have actually been reported in the literature. To the best of our knowledge, this is the first report of familial multiple isolated EICs that is compatible with X-linked inheritance. The present report unifies the description of a detailed multi-generational family history for provision of the most accurate recurrence risk information.

2788T
Intragenic deletion of NPAS3 in a child with developmental delay. C. Armour1,2,3, E. Baxter1, J. McGowan-Jordan1,2,3. 1) Eastern Ontario Regional Genetics Program, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON; 3) Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada.

NPAS3 (neuronal PAS domain-containing protein 3) is a basic-helix-loop helix transcription factor expressed in the developing brain and implicated in neurodevelopment as it is thought to function in neuron production and maturation. A number of reports have suggested that disruption of NPAS3 may result in both cognitive impairment and psychiatric illness. Interestingly, this gene has also been shown to contain the largest number of human-accelerated regulatory sequences (non-coding conserved elements that are unique to humans) and thus postulated to have contributed to human brain evolution. This report describes an 8 year old female with developmental delay and an intragenic deletion of NPAS3. She walked at 2.5 years and began school at 4 years of age and has additional delays in fine and gross motor skills. She currently attends school in a modified class and overall abilities are currently estimated to be 2 to 3 years behind. Additional medical issues include: daytime intermittent enuresis, an asymptomatic spinal syrinx, obesity (BMI of 34) and an astigmatism. Physical exam reveals macrocephaly of approximately 3 to 4 standard deviations (SD) above the mean and height at approximately 2 standard deviations from the mean. She has a round face with full cheeks, mild ptosis and down slanting palpebral fissures. Family history is notable for a maternal history of learning disability requiring special education. The father has macrocephaly, approximately 3 SD above the mean, and has a historical diagnosis of a Fetal Alcohol Spectrum disorder. High density oligonucleotide array in the proband revealed a 0.11Mb intragenic deletion at 14q13.1 involving only NPAS3. Prader-Willi syndrome/1P36 studies and PTDN sequencing were normal. Follow-up studies performed so far indicate that the deletion was maternally inherited, correlating with the mother's own reported learning disability. Family studies are ongoing. This case supports a role for NPAS3 in neurodevelopment and abnormal NPAS3 function may be involved in other cases of non-specific cognitive impairment.

2789F
Dyggve-Melchior-Clausen syndrome In Three Generation. B. Bozorgmehr, A. Kammiejad. Dept Clinical Genetics, K-N Pathology & Genetics Ctr, Tehran, Iran.

Dyggve-Melchior-Clausen(DMC) syndrome is a rare autosomal recessive spondyloepimetaphyseal dysplasia. The main features are short trunk dwarfism, short limbs with intellectual disabilities. Radiological findings are: abnormalities of spine, epiphyses and metaphyses. We are reporting a 15-year-old boy with Afghan origin with short trunk dwarfism, short limbs, characteristic radiological findings and severe mental retardation who had three affected maternal uncles and also his mother had one affected maternal uncle. His parents were first cousin and the others, were from one little village so their pedigree showed a pseudo X-linked recessive pattern.

2790W
Novel de novo SPOCK1 Mutation in a Proband With developmental delay, microcephaly and agenesis of corpus callosum. R. Dharmia1, J. Graham2, E. Thorland3, S. Kimman1. 1) Mayo clinic, Rochester, MN; 2) Cedars-Sinai Medical Center, Los Angeles, CA.

Introduction: Whole exome sequencing of an affected patient and unaffected parents has now made it possible to identify novel de novo mutations in genes possibly linked to human disease. Here we describe a patient with developmental delay and partial agenesis of corpus callosum with a novel de novo SPOCK1 variant. Mutations in SPOCK1 have not been previously described to cause a human phenotype. Case report: The proband's phenotype included severe global developmental delay with dystonia, structural brain abnormality with partial agenesis of corpus callosum, microcephaly and atrial septal defect with aberrant subclavian artery. Extensive chromosomal, genomic and metabolic studies were unrevealing. At age of 13 years, exome sequencing was performed on the patient and her parents at GeneDX, Gaithersburg, MD, and a de novo novel missense mutation was identified in SPOCK1 (coding for Testican-1) on chromosome 5q31: c.239A>T (p.D80V).

This mutation affects a highly evolutionarily conserved area of the gene, replacing a polar aspartic acid with hydrophobic nonpolar valine that changes the chemical properties of the protein product, likely representing a pathogenic variant. Online prediction programs SIFT (http://sift.jcvi.org/) and mutation taster (http://dorcharlie.de/MutationTaster/index.html) suggest this variant is a deleterious mutation. Previous microdeletions of 5q31 including SPOCK1 have suggested genes on 5q31 as candidates for intellectual disability. No mutations or variants in other genes potentially linked to the phenotype were identified. Conclusion: Testicans are proteoglycans belonging to the BM-40/SPARC/osteonectin family of extracellular calcium-binding proteins. Testican-1 is encoded by the SPOCK1 gene and in mouse models has been shown to be strongly expressed in the brain and modulate important steps in neurogenesis. In humans the expression is more widespread and also seen in heart, blood and cartilage. We hypothesize that because this gene function is critical in neurogenesis, mutations can potentially lead to a phenotype with developmental delay and microcephaly. Further functional studies are under way to confirm these findings, potentially making this the first reported case of human disease related to SPOCK1.
2791T FOXP1 mutations cause intellectual disability and a recognizable phenotype. M.F. Hunter1,2, A.K. Le Feuvre2, S. Taylor3, N.H. Malek4, D. Horn5, C.W. Carpenter6, O.A. Abdul-Rahman7, S. O'Donnell8, T. Burgess8, M. Shaw9, J. Geck2, N. Bain10, K. Fagan11,2. 1) Hunter Genetics, Waratah, Australia; 2) John Hunter Children’s Hospital, Newcastle, NSW, Australia; 3) Core Interventions Occupational Therapy Services, Gosford, NSW, Australia; 4) Psychiatry Pathologist, Lisarow NSW; 5) Institute of Medical Genetics, Charité, University of Berlin, Berlin, Germany; 6) Department of Dermatology, Emory University, Atlanta, GA, USA; 7) Department of Pediatrics, University of Mississippi Medical Center, Jackson, MS, USA; 8) VCGS Pathology, Melbourne, and Department of Pathology, Rush University Medical Center, SA, Australia; 9) Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia; 10) Faculty of Health, University of Newcastle, Newcastle, Australia.

Mutations in FOXP1, located at 3p13, have been reported in patients with global developmental delay (GDD), intellectual disability (ID) and speech defects. Mutations in FOXP2, located at 3q31, are well known to cause developmental speech and language disorders, particularly developmental verbal dyspraxia. FOXP2 has been shown to work co-operatively with FOXP1 in human development. An overlap in FOX1P1 and FOX2 expression, both in songbirds and the human fetal brain have suggested that FOXP1 may also have a role in speech and language disorders. We report a male child with a 0.19Mb intragenic deletion that is predicted to result in haploinsufficiency of FOXP1. Review of our patient and others reported in the literature reveals an emerging phenotype of GDD/ID with moderate to severe speech delay where expressive speech is most severely affected. Facial features include a broad forehead, down-slanting palpebral fissures, and a short nose with broad tip, relative or true macrocephaly, a frontal hair upsweep and persistent fetal finger pads. Autistic traits and other behavioral problems are likely to be associated. Congenital malformations may also be associated with FOXP1 haploinsufficiency.

2794T Microarray analysis of a de novo microdeletion involving the Van der Woude Syndrome locus. E.C. Tan1,2, E.C.P. Lim1, S.T. Lee1. 1) KK Research Centre, KK Women’s & Children’s Hospital, Singapore; 2) Department of Plastic, Reconstructive & Aesthetic Surgery, Singapore General Hospital, Singapore.

Van der Woude syndrome is the most common among syndromes which include cleft lip and/or cleft palate as one of the presentations. This autosomal dominant disorder is usually caused by mutations in the interferon regulatory factor 6 (IRF6) gene, with a few cases of small deletions involving coding sequences in the gene. For larger deletions, there were only four reports of microdeletions in this chromosomal region, one was microscopic while the other three involved microsatellite markers outside the IRF6 gene region. We describe the detection of an interstitial deletion of approximately 2.3Mb within the 1q32.2 region in a male Chinese patient presenting with typical VWS features using the Affymetrix Human SNP 6.0 Array. Consistent with his negative family history, the same deletion was not present in either parent and his two siblings (a younger sister and a younger brother) were also phenotypically normal. Using microsatellite markers, it had been established that the deletion occurred on the paternal chromosome. Although several known genes (including 11 OMM genes and several micro RNAs) are deleted, this patient has no other abnormality apart from the orofacial presentations typical of VWS. He has been followed up since birth and there has been no clinically significant abnormality in any other organ system. Other than IRF6, the genes which are deleted in this patient appear to be insensitive to copy number and haploinsufficiency. The deletion in this patient overlaps with another VWS patient which is the only published case with array-CGH data. The female patient is also from this geographical region and was reported to have a 3.4Mb deletion. However, she has dysmorphism, some growth retardation and some growth retardation in addition to the VWS phenotype. We will present a comparison of the different breakpoints and genes deleted in these two patients and discuss the implications.

2792F CTNNB1 mutation in siblings with intellectual disability, spasticity and microcephaly. D. Tegay1,2, J.Gecz3,4. 1) Department of Medicine, NYIT College of Osteopathic Medicine, Old Westbury, NY; 2) Division of Genetics, Nassau University Medical Center, East Meadow, NY.

De novo heterozygous loss-of-function mutations in the beta catenin-associated protein, beta-catenin (CTNNB1), have recently been identified through whole-exome sequencing in 3 unrelated individuals sharing similar features of intellectual disability, absent or limited speech, microcephaly and spasticity (de Ligt et al., 2012). Beta-catenin is a key downstream component of the canonical Wnt signaling pathway and acts as a negative regulator of centrosome cohesion. While somatic gain-of-function activating mutations predicted oncogenic activity, animal models suggest germline loss-of-function mutations affect neuronal, particularly cerebellar, development and maturation. We report the first case of siblings, both presenting with variable degrees of unexplained intellectual disability, microcephaly, spasticity and non-specific dysmorphic features, in whom clinical whole-exome sequencing identified a common heterozygous CTNBN1 nonsense mutation (S681X) not detected in either parental sample, suggesting parental germline mosaicism as the likely molecular mechanism. These cases help further delineate and expand the phenotype of CTNBN1 mutation.

2793W Molecular genetic characterization of INSR in a family with Rabson-Mendenhall syndrome. F. Kantheti1, A. Raman2, J. Bhattacharyya3, J. Conneally2, A. Sumathi1,4, P. Kantheti1,2, M. Mendenhall1,2. 1) Department of Medicine, NYIT College of Osteopathic Medicine, Old Westbury, NY; 2) Division of Genetics, Nassau University Medical Center, East Meadow, NY.

Rabson-Mendenhall syndrome (RMS) is a rare autosomal recessive disorder caused by mutations in the insulin receptor gene (INSR). Sequence analysis of INSR in a consanguineous family with all the offspring affected revealed that the two surviving female index cases have a homozygous nucleotide substitution c.768G>C at the 220th codon of INSR predicted to change glycine to arginine within the insulin binding domain of the alpha subunit. The mutation does not appear to affect the transcription of the gene or the binding of insulin to the receptor. Haplotypes inferred from single nucleotide polymorphisms indicated that both siblings inherited identical haplotypes linked to the same causative mutation. While the other three involved microsatellite markers from the parents who are heterozygous (His 1085 T/C). This polymorphism may possibly lead to the disruption of an E-box motif (CAGGTG) and/or to the disruption of a potential epigenetic signature. This may have relevance to the insulin resistance commonly seen in RMS patients.

2795F A novel TPM3 gene mutation with infantile nemaline myopathy. S.O. NAM1,1, J.H. SHIN2,3, Y.J. LEE4,1. 1) Pediatrics, Pusan National University Children’s Hospital, Yangsan, Gyeongnam, South Korea; 2) Neurology, Pusan National University Children’s Hospital, Yangsan, Gyeongnam, South Korea.

A twelve month-old girl was transferred with respiratory failure. She had no specific perinatal problem with gestational age of 38 weeks and 4 days and birth weight of 3.680g. At one month of age, her parents found that her muscle power was decreased. Her parents visited a nearby rehabilitation center and the patients had been treated with physical therapy. At 11 months of age, she had caught flu-like symptoms lasting for 2 weeks and progressed to abrupt cyanosis and respiratory failure and visited emergency room. She was admitted to intensive care unit on ventilation therapy after cardiopulmonary resuscitation for 8 minutes. She was tried to weaned off her ventilator but failed several times, so she was transferred to the pediatric neurologic clinic of our hospital for the evaluation of underlying disease. There was a family history of motor delay of her mother and her grandfather in her mother’s side grown out at adulthood. Physical examination showed elongated face with high arched palate, but had no dysmorphism or visceromegaly. Theophylline examination showed alert mentality and normal cranial nerve function. However, there was marked gross motor delay without involvement of fine motor, language or cognition. Her muscle tone was markedly decreased with frog leg posture, positive head lag sign, and decreased deep tendon reflex. Her laboratory findings including serum creatine kinase and metabolic study were normal. Her brain MRI, electroencephalography, EMG/NCV and cardiac echocardiography were normal. Genetic evaluation including SMN1 gene, DMPK gene, fragile X PCR, and array CGH were normal. The embryonic findings of muscle biopsy revealed selective type 1 fiber atrophy with accumulation of suspicious dense material in modified Gomori trichrome stain. Electromicroscopic findings showed severe loss of myofibrillar structure within nemeal rod accumulation within atrophic fiber. Whole exome sequencing was performed. The results showed 151 variants in genes previously reported to cause neuromuscular disorders. The number of candidate variants was reduced to 22 when screened by normal Korean SNP database. We found a novel missense heterozygous mutation in exon 1 of TPM3 gene (c.327T>A, p.Met115Lys) close to the known pathogenic mutation, compatible with the diagnosis of nemaline myopathy.

Rhabdomyolysis is a condition of rapid breakdown of damaged skeletal muscle triggered from various causes including exercise, hyperthermia, and anesthetic agents. Patients with inherited muscle disorders such as muscular dystrophies are at great risk for malignant hyperthermia and rhabdomyolysis upon using different types of anesthetic agents. This condition can be life-threatening and usage of those medications should be avoided for those patients. We report an infant with concurrent diagnosis of Freeman-Sheldon syndrome (FSS) and Duchenne or Becker muscular dystrophies (DMD/BMD), who was admitted to our pediatric ICU for cerebral infarction and elevated CPK after general anesthesia. The patient was a 3-month old male born at full term without complications to non-consanguineous parents from the Dominican Republic with a benign family history. At birth, he had left undescended testis, right inguinal hernia, and developmental dysplasia of the hips. Several days after the cryptorchidism and hernia surgery, he developed a fever, abnormal eye movements, facial nerve palsy and hypertonia. Brain MRI showed watershed cerebral infarction and his CPK was elevated above 45,000 U/L. Physical examination revealed deep-set eyes, low-set ears, high-arched palate, H-shaped cutaneous dimpling of the chin, cortical thumbs, and contractures of wrist and ankle joints. The physical findings supported the diagnosis of FSS, an autosomal dominant disorder due to mutations in MYH3 gene and characterized by typical facial features, distal arthropathy and susceptibility to malignant hyperthermia with rhabdomyolysis following exposure to inhalational anesthetics. Array CGH was done to look for contiguous gene deletion and incidentally identified a de novo 74.5-kb deletion on chromosome Xp21.1. The deletion included exons 45-47 of the DMD gene, which was predicted to cause DMD or BMD. This is the first reported case of two separate congenital myopathic diseases arising de novo in one individual. We speculate that the stroke episode was a complication of malignant hyperthermia with rhabdomyolysis triggered by sevoflurane for general anesthesia. Muscle rigidity has been reported in both DMD and BMD patients, adding to the diagnostic challenge in infants. We believe this is likely to be a result of the two separate myopathic processes in this patient.

2797T Whole-exome sequencing identifies a homozygous mutation in the SPG11 gene in patients with Spastic Paraplegia. V. Adir1, A. Shalata2, E. Shahak3, Z.U. Borochowitz1. 1) Simon Winter Institute for Human Genetics, Bnai Zion Medical Ctr., Haifa, Israel; 2) Genetic Institute, Ziv Medical Ctr., Zefat, Israel.

Hereditary spastic paraplegias (HSP) are neurodegenerative diseases which include a heterogeneous group of neurodegenerative diseases. HSP is mainly characterized by lower limb spasticity associated with additional neurological signs in the complicated forms. At least 52 loci and 31 causative genes have been identified. Although mutations in the SPAST gene explain approximately 40% of the pure autosomal dominant forms, molecular diagnosis is challenging for the sporadic and recessive forms, which are often complicated and clinically overlap with a broad number of movement disorders. In this study whole-exome sequencing was performed to reveal the genetic cause for the disease in a large family with consanguineous marriage in northern Israel. We sequenced two first cousins patients and the father of one of them. Exome sequencing revealed that both patients carry a novel homozygous nonsense mutation in the SPG11 gene (c.4339C>T; p.Q1447X). The presence of the mutation was confirmed by Sanger sequencing. We developed a PCR RFLP test that would enable easy analysis of the c.4339C>T mutation in the SPG11 gene. The mutation was not found in 100 control chromosomes from the same ethnicity. According to previous studies, the SPG11 encodes the spastatin protein which is expressed ubiquitously in all tissues. The function of spastatin remains unknown, but the fact that it is highly conserved among species suggests an essential biological function. Further studies which will reveal the normal function of spastatin will contribute to the understanding of the pathology of the disease, and better therapy for the patients.


Renal hypouricemia is a common inherited disease characterized by low serum uric acid (SUA) levels. It is associated with severe complications such as urolithiasis and exercise-induced acute renal failure. We have previously reported that urate transporter 1 (URAT1/SLC22A12) and glucose transporter 9 (GLUT9/SLC2A9) are involved in renal hypouricemia type 1 (RHUC1) and renal hypouricemia type 2 (RHUC2), respectively. In the series of experiments, two families have been revealed to have RHUC2 due to GLUT9/SLC2A9 missense mutations R198C or R380W. However, there are no reports of other RHUC2 families or patients with these pathogenic mutations until now. Thus, we performed mutational analysis of GLUT9/SLC2A9 exon 6 (for R198C) and exon 10 (for R380W) in another 50 hypouricemia patients to find other cases of RHUC2. Patients were collected out of more than 2000 samples from J-MICC Study (the Japan Multi-Institutional Collaborative Study). In this study, one novel hypouricemia male patient with heterogeneous RHUC2 mutation R380W was identified. His SUA level was 2.6 mg/dL, which is similar to that of our previous case (SUA level: 2.7 mg/dL). Mutation sites in GLUT9/SLC2A9 (R380W and R198C) locate in highly conserved amino acid motifs in ‘sugar transport proteins signatures’, which is observed in GLUT family transporters. As for glucose transporter 1 (GLUT1) deficiency syndrome which is caused by the corresponding mutations in human GLUT1 (R333W and R198C), mutation sites in GLUT9/SLC2A9 are not important determinant of membrane topology of GLUT1. The same is probably true in GLUT9 on the basis of membrane topology. In oocyte expression study, their mutants showed markedly reduced urate transport, which would be the result of loss of positive charges of those amino acid motifs. The identification of the new RHUC patient could help to expand the understanding of RHUC pathogenesis and suggest that these GLUT9/SLC2A9 mutations cause renal hypouricemia by their decreased urate reabsorption on both sides of the renal proximal tubules. In addition, these findings also enable us to propose a physiological model of the renal urate reabsorption and can be a promising therapeutic target for hyperuricemia, gout and associated diseases, such as cardiovascular diseases, cerebrovascular diseases and renal failure.

2799W The manifestation of Mowat-Wilson syndrome in adult identical twins. S. Ramanathan1, S.A. Ashwal2, R.D. Clark1. 1) Pediatric Genetics, Loma Linda University, San Bernardino, CA; 2) Pediatric Neurology, Loma Linda University, San Bernardino, CA.

Mowat-Wilson syndrome (MWS) is characterized by typical dysmorphic facial features, congenital anomalies, including Hirschsprung disease, and moderate to severe intellectual disabilities. Here we report the case of the twins who were diagnosed with MWS at 2 years of age. Their muscular hypotonia, feeding difficulty, and developmental delay was initially attributed to the twin gestation and lower birth weight (4lb 11oz, 5lb 2 oz) compared to their siblings. At 14 months, the twins were microcephalic and were not rolling over. MRI of the brain showed abnormality in the cerebellum. Both have epilepsy, which is well-controlled in one. Both have chronic constipation. As adults, they are non-verbal. They babble, follow one-step commands and have some receptive language. They are not toilet-trained. Neither is able to chew food. They need to be restrained at meal-time, as they bat away at spoons. Both have bruxism and bite their wrists. Menarche was at 16 and 18 years and menses are irregular. The twins have had normal echocardiograms, and are awaiting skeletal biopsies to rule out Hirschsprung disease. They are being followed by a rheumatologist for swelling of the interphalangeal joints, which has been reported in MWS, but is not well-characterized. Radiographs on the twins do show soft tissue, but no bony changes. Our patients illustrate the challenge of retaining a medical home while having access to the larger medical data on adults with MWS. The authors and the parents of these patients welcome contact with other families who have adults with MWS to collect data on the natural history of this disorder.

Charcot-Marie-Tooth (CMT) disease is a heterogeneous group of inherited peripheral neuropathies. CMT type 4 is characterized by autosomal recessive inheritance. CMT4B1 due to MTMR2 gene mutations is a rare subtype of CMT4 with reported association of facial weakness, vocal cord paralyses, chest deformities, and claw hands. We report the unusual occurrence of recurrent optic neuritis and cervical cord Schwanomna in a male individual with confirmed CMT4B1 disease. This 28-year-old gentleman presented with bilateral, symmetrical limb weakness that started distally and progressed very slowly since the age of 4 years and became wheel chair dependent at the age of 22. Dysphonia with vocal cord paralysis was apparent as he grew. He developed 3 episodes of optic neuritis resulting in a reduced visual acuity. At the age of 27 years, he presented with symptoms suggestive of cervical cord compression with diagnosis of C2/C3 spinal cord Schwanomna. Physical examination was remarkable for facial nerve weakness, chest deformity and claw hands. A positive family history with two affected siblings and the presence of consanguinity suggest autosomal recessive inheritance. The clinical features were consistent with CMT4B1 and MTMR2 gene sequencing revealed a novel nonsense homozygous mutation c.1768C>T (p.Gln590*). The rare association of optic neuritis or Schwanomna with genetically confirmed CMT1A has been individually observed but never with recessive CMT. To the best of our knowledge, the occurrence of recurrent optic neuritis and cervical cord Schwanomna in the same patient has never been reported with any forms of CMT including CMT4B1.

2801F Novel homozygous missense mutation in the matrix metalloproteinase 2 (MMP2) catalytic domain leading to protein loss-of-function in two Italian sibs in the spectrum of Torg-Winchester syndrome. J. Azzollini1, D. Rovina1, C. Gervasin1, I. Parenti1, A. Fratoni1, L. Pietrogrande1, L. Larizza1. 1) Medical Genetics, Department of Health Sciences, University of Milan, Milan, Italy; 2) Pediatrics Unit, G. Salvini Hospital, Rho, Italy; 3) Orthopaedic Division, Department of Health Sciences, University of Milan, Italy.

Torg-Winchester Syndrome (TWS, OMIM #259600) is an extremely rare autosomal recessive disorder characterized by multicentric osteolysis, especially involving hands and feet, subcutaneous nodules, arthropathy with progressive joint contractures and other features such as short stature, hyperpigmented skin lesions, coarse face, corneal opacities and gum hyper trophy. Mutations of MMP2 gene (16q13), encoding Matrix Metalloproteinase 2, have been associated with TWS: only 8 mutations, 3 truncating, 3 missense, 1 splice, 1 in-frame deletion have been described so far in 8 families from different geographic areas. We evaluated two Italian siblings (43°, 37°), born to healthy consanguineous parents, who started to display, between ages 2 and 6, conspicuous osteolysis and progressive generalized subcutaneous nodules followed, during adolescence, by tendon retractions, progressive upper limb arthropathies, toes osteolysis and pigmented fibrous skin lesions. The female also suffered from recurrent bilateral eye pterygium after age 30. Based on the clinical presentation and the STRs genotype shared by the sibs at 16q12.2-q21, we performed MMP2 mutation screening on their blood DNA by direct sequencing. Both siblings were found to carry a novel homozygous missense mutation in exon 8, c.1228G>C p.G410R, affecting a highly conserved aminoacid residue within MMP2 catalytic domain. The same mutation was not detected in 260 healthy individuals and is predicted pathological by several bioinformatic tools (Mutation Taster, PolyPhen-2, PMut, Mutation Assessor). Functional assessment was carried out by site-directed mutagenesis on the wild-type MMP2 sequence cloned into an expression vector, trassection of both wild-type and mutated MMP2 into HEK cell line and gelatin-zymography on the conditioned conditioned media. As compared to the wild-type MMP2, G410R-MMP2 showed a complete loss of gelatinolytic activity, thus confirming the mutation impact on protein function. Out of 8 described MMP2 mutations, map to exon 8, in the protein catalytic domain (p.E404K, p.V400del, p.G406D), and are associated with Winchester phenotype, mainly characterized by the absence of subcutaneous nodules. p.G410R represents the first inactivating mis sense mutation, within MMP2 catalytic domain, associated with Torg phenotype, thus reinforcing the complexity of genotype-phenotype correlation in TWS.

2802W A patient with Simpson-Golabi-Behmel syndrome, Biliary Cirrhosis and successful Liver Transplantation. B. Demeur1, G. Guillaume Jaredzak1, M. Girard1, A. Mellos2, D. Djedjé3, C. Charod1, A. Varentcherm1, M.P. Moizard1, J. Gondry1, H. Sevestre1, M. Mathieu-Dramard1, F. Lacaille2. 1) Genetics Dept, Hopital nord, CHU, Amiens, France; 2) Paediatric Hepato-gastroenterology-Nutrition unit, Necker-Enfants-malades Hospital, France; 3) Paediatric hepatogastroenterology-Nutrition unit, C.H.U Amiens, France; 4) Paediatric surgical unit, Necker-Enfants-malades Hospital, France; 5) Service de génétique, INSERM U930, CHRU Tours, France; 6) Prenatal diagnosis unit, Camille Desmoulins maternity hospital / Jules Verne University, C.H.U Amiens, France; 7) Pathology department, C.H.U Amiens, France.

Simpson-Golabi-Behmel syndrome (SGBS)-OMIM #312870, -reported by Simpson et al in 1975 is a rare X-linked inherited overgrowth syndrome caused by a loss-of-function mutation in the GPC3 (SGBS type 1) or GPC4 gene. Patients present a variable phenotype with pre- and post-natal macrosomia, distinctive facial dysmorphism, organomegaly, and numerous congenital anomalies, i.e: diaphragmatic hernia, heart or renal defect, genito-urinary tract or gastrointestinal malformations, skeletal or hand abnormalities. Intellectual disability is not constant. About 10% of patients develop embryonic tumors in early childhood. A single case of biliary disease (choledochal cyst) has been reported. GPC3 is localised on Xq26 and encodes for Glypican-3, a glycosylphosphatidylinositol-linked cell surface heparan sulfate proteoglycan, that belongs to the family of glycipicans. This report concerns a male infant with Simpson-Golabi-Behmel syndrome type 1, carrier of a GPC3 mutation. He had neonatal liver disease, and developed early on biliary cirrhosis. Liver transplantation was discussed, considering the risks of cancer and intellectual disability, and performed successfully when the child was 19 months of age. A hypothesis on the role of GPC3 in this patient’s liver disease is proposed.


Waardenburg Syndrome (WS) is characterized by sensorineural hearing loss and pigmentedary differences of the hair, skin and iris with both phenotypic variability, as well as genetic heterogeneity leading to four subtypes. The absence or presence of dystopia canthus (lateral displacement of the inner canthi) historically has directed clinical testing for two of the four subtypes, as WS type I (presence of dystopia canthus) is associated with mutations in PAX3, whereas WS type II (absence of dystopia canthus) is associated with mutations in MITF. We report on a 5-month old proband referred for bilateral sensorineural hearing loss whose physical exam was notable for dystopia canthus by the presence of a Waardenburg-index of 2.04 (normal being less than 1.95) and whose father was also affected, suggesting dominant inheritance. PAX3 sequencing and deletion analysis on the proband and was negative. Additional samples from the proband, unaffected sibling, father and mother were obtained and exome sequencing was performed by BGI @CHOP Genome Center on an Illumina HiSeq 2000. Bioinformatic analysis of potentially pathogenic sequence variants shared in the proband and father was completed and led to a single candidate variant as the cause of the phenotype - an MITF p.R255X nonsense mutation previously associated with WS. Initially, this variant was not identified at a read depth coverage of 10x, but was revealed when the read depth coverage was relaxed to 8x. Sanger sequencing was performed to validate the finding and it was present in the proband and father, as well as in the presumed unaffected sibling and was not present in the unaffected mother. This case report of using exome sequencing to aid in the diagnosis for a family with bilateral sensorineural hearing loss highlights the importance of: 1) BGI’s ability to identify genetic alterations that would otherwise not have been tested for clinically; 2) amending historically established phenotypes to incorporate a larger clinical spectrum; 3) using lower coverage cuts off in an attempt to decrease false negatives; and 4) how the result of this testing can unearth reduced penetrance in presumed ‘unaffected’ family members.
Use of quantitative ultrasound for tibial dysplasia in neurofibromatosis type 1.}

D.A. Stevenson, H. Slater, H. Hanson, A. Stevens, J.C. Carey, D.H. Viskochil. Dept Pediatrics, Univ Utah, Salt Lake City, UT.

Neurofibromatosis type 1 (NF1) is a common autosomal dominant genetic disorder with distinct skeletal manifestations. In particular, tibial dysplasia is a difficult-to-treat morbid manifestation. Tibial dysplasia typically presents with unilateral anterolateral bowing of the tibia with cortical thickening and medullary canal narrowing, with subsequent fracture and non-union. In infancy, physiologic bowing of the lower leg can be confused with pathologic tibial dysplasia in NF1, and early intervention with bracing is thought to improve outcomes. Quantitative ultrasound (QUS) measures speed of sound, avoids radiation, and is reported to be predictive of clinical fracture in the general population. Our aim was to use QUS of the affected and unaffected tibias of individuals with NF1 who presented with suspected tibial dysplasia prior to fracture and pseudarthrosis. Bone quality was assessed on both tibia (the non-bowed and bowed tibia) using the Sunlight Omnisense 7000P scanner (Sunlight Medical, Israel) to measure the speed of sound (SOS) at the mid-shaft. The probe is moved in a sweeping motion, laterally and medi ally, and three to five consecutive measurement cycles are performed, after which the SOS (m/s) is determined and z-scores generated using cross-sectional reference data of the same sex and age provided with the machine. A total of 14 individuals with NF1 with unilateral tibial bowing without fracture/pseudarthrosis were enrolled and both tibia were tested using QUS. The mean SOS z-score for the affected tibia was -2.9. The mean difference in SOS z-score was -2.5 when comparing the bowed tibia versus the individual’s contralateral unaffected tibia. One infant with mild anterolateral bowing, radiographically showed no cortical thickening, and the z-score of the bowed bone had a SOS z-score +2.5 difference compared to the other tibia suggesting physiologic bowing rather than pathologic tibial dysplasia. These data show that dysplastic tibia in NF1 prior to fracture and non-union have abnormal bone quality with significant decreases in SOS. These data also suggest that QUS is sensitive to distinguish dysplastic bowing vs. physiologic bowing in infancy in NF1. QUS is an attractive tool as a quantitative outcome measure for future trials aimed at improving tibial bowing to prevent fracture, and potentially for an aid in diagnosis and clinical management in NF1.

New syndrome of ectrodactyly and lethal pulmonary acinar dysplasia associated with homozygous FGFR2 mutation identified by exome sequencing.

C.P. Barnett, H.S. Nataren, M. Köngler-Hoffmann, Q. Schwarz, D.L. Bruno, J. Lipsett, A.J. McPhee, A.W. Schreiber, J.H. Feng, C.N. Hahn, H.S. Scott, C.P. Barnett, C.N. Hahn, D.L. Bruno, A.J. McPhee, A.W. Schreiber, J.H. Feng, C.N. Hahn, H.S. Scott, 1) SA Clinical Genetics, Women’s and Children’s Hospital/SA Pathology, North Adelaide, South Australia, Australia; 2) Department of Molecular Pathology, Centre for Cancer Biology, SA Pathology, Adelaide, SA, Australia; 3) School of Molecular and Biomedical Science, University of Adelaide, SA, Australia; 4) School of Medicine, University of Adelaide, SA, Australia; 5) Cytogenetics Laboratory, Murdoch Children’s Research Institute, Royal Children’s Hospital, Parkville, Australia; 6) ACRF Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology; 7) Department of Neonatal Medicine, Women’s And Children’s Hospital, North Adelaide, SA, Australia; 8) Department of Anatomical Pathology, Women’s and Children’s Hospital/SA Pathology, North Adelaide SA, Australia.

A female infant from a consanguineous union was born at term but died of respiratory failure on day 2. An autopsy revealed severe pulmonary hypoplasia due to acinar dysplasia/congenital pulmonary airway malformation (CPAM) type 6. The infant also had ectrodactyly involving the hands and feet. Ectrodactyly is genetically heterogeneous and acinar dysplasia is a rare lethal congenital lung lesion of unknown etiology. Recessive inheritance has been suggested. SNP array and whole exome sequencing of this infant’s DNA identified 22 homozygous candidate causative sequence variants (SV) after filtering against dbSNP and numerous published in-house exomes. Furthermore, analyses based on prediction of pathogenicity, protein conservation and expression (especially embryonic lung and limb), highlighted one SV in the FGFR2 gene, homozygous missense variant R255Q. Numerous germline FGFR2 mutations have been described in association with human disease, all heterozygous and mostly de novo autosomal dominant mutations associated with craniosynostosis syndromes. Involved one SV in the FGFR2 gene, homozygous missense variant R255Q. Numerous germline FGFR2 mutations have been described in association with human disease, all heterozygous and mostly de novo autosomal dominant mutations associated with craniosynostosis syndromes. Involved one SV in the FGFR2 gene, homozygous missense variant R255Q. Numerous germline FGFR2 mutations have been described in association with human disease, all heterozygous and mostly de novo autosomal dominant mutations associated with craniosynostosis syndromes. Involved one SV in the FGFR2 gene, homozygous missense variant R255Q. Numerous germline FGFR2 mutations have been described in association with human disease, all heterozygous and mostly de novo autosomal dominant mutations associated with craniosynostosis syndromes. Involved one SV in the FGFR2 gene, homozygous missense variant R255Q.
2808T
A GC polymorphism associated with serum 25(OH)D level is the risk for hip fracture in Japanese patients with rheumatoid arthritis. S. YOSHIDA1, K. IKARI1, T. FURUYA1, Y. TOYAMA2, A. TANIGUCHI1, H. YAMA-NAKA1, S. MOMOHARA2. 1) Institute of Rheumatology, Tokyo Women's Medical University, Shinjuku, Tokyo, Japan; 2) Department of Orthopaedic Surgery, School of Medicine, Keio University, Tokyo, Japan.

Background: Vitamins are important for the maintenance of the musculoskeletal system, is positively associated with muscle strength and physical performance, and is inversely associated with falls and fracture risk. The major determinants of serum 25-hydroxyvitamin D [25(OH)D] concentration are sunlight exposure on maxillary and mandibular arc measurements is primarily on cases with a FASD. We did not find a significant difference in the arc or the ratio in measurements in male and female groups but not the ratio. The trend controls. With a few gender-based differences, the arc measurements were defined as less than 10th centile, individually on arc measurements in the controls. We evaluated the effect of microcephaly, short stature, and low weight, respectively. The ratio was 0.9643 and 0.9676 for males and females respectively.

We investigated the genetic variants associated with serum vitamin D concentration in the occurrence of hip fracture in Japanese patients with rheumatoid arthritis (RA). Methods: DNA samples of 2068 Japanese patients with RA were obtained from the Institute of Rheumatology Rheumatoid Arthritis cohort study (IORRA) DNA collection. Serum 25(OH)D concentration was measured in 932 of 2068 patients in the spring of 2011. Five of the single nucleotide polymorphisms (SNPs) reported in the recent studies were genotyped: rs2286279, in GC; rs3829251, in NADSYN1; rs12785878, rs1790349, near and in DHCRT1, and rs10741657, near CYP2R1. The occurrence of hip fractures was determined from the responses to a patient questionnaire every 6 months from October 2000 to October 2010. After confirmed by review of medical records and radiographs, 39 hip fractures in 39 patients were included into this study. Multivariate linear regression analyses adjusted for the non-genetic factors were performed to investigate the association between serum 25(OH)D concentration and hip fracture risk in patients. The SNP showing a significant association with serum 25(OH)D concentration was included into the following prediction analysis for the occurrence of hip fracture by using a multivariate Cox proportional hazards regression model. Results: Multivariate Cox proportional hazards regression model indicated that rs2286279 in GC was significantly associated with the occurrence of hip fracture [HR (95% CI): 2.03 (1.03 to 5.94); P=0.042]. Conclusion: Our data indicated that a GC polymorphism is the risk for hip fracture in Japanese patients with RA.

Fetal alcohol syndrome and assessment of maxillary and mandibular arc measurements. K. Abeli1, W. May1, P. May2, W. Kahberg1, G. Hoyme1, O. Abdul-Rahman1. 1) University of Mississippi Medical Center, Jackson, MS; 2) University of New Mexico, Albuquerque, NM; 3) University of North Carolina, Chapel Hill, NC; 4) University of South Dakota, Sioux Falls, SD.

Fetal alcohol spectrum disorders (FASD) are a range of physical differences and neurologic deficits due to prenatal alcohol exposure. Diagnosis evaluates characteristics like growth deficiency, microcephaly, and cardiac facial features, and other features like midface hypoplasia are also seen. While traditionally judged subjectively, objective measurements of maxillary and mandibular arcs and the ratio between them may be useful in assessing midface hypoplasia. The Fetal Alcohol Syndrome Epidemiologic Research (FASER) project outlines morphologic assessments of children for FAS. Data are collected during in-school screenings of first-graders for height, weight, and head circumference. Cases are selected by measurements below the 10th percentile, and matched with normal controls. All subjects are evaluated by a dysmorphologist for FASD physical features. Those with possible features undergo neuropsychiatric evaluation and maternal interview for a diagnosis of FAS, partial FAS, alcohol-related neurodevelopmental disorder (ARND), or non-FAS. Using the FASER database, we investigated the size of the maxillary and mandibular arcs and the arc ratio in respect to FAS diagnosis.

We established normative values for the maxillary arc, mandibular arc, and the maxillary-to-mandibular arc ratio for males and females. In our control group of 545 males and 436 females, the mean maxillary and mandibular arcs were 24.98/24.52 cm and 25.91/22.35 cm respectively. The ratio was 0.9643 and 0.9676 for males and females respectively. We evaluated the effect of microcephaly, short stature, and low weight, defined as less than 10th centile, individually on arc measurements in the controls. With a few gender-based differences, the arc measurements were reduced significantly but the ratio did not differ. We compared our controls to 138 males and 135 females FASD cases. We noted a significant difference in arc measurements in male and female groups but not the ratio. The trend was greatest for diagnosis of FAS but also seen in partial FAS and ARND. We compared non-FAS controls with reduced growth parameters to similar cases with a FASD. We did not find a significant difference in the arc or the ratio measurements. Therefore, we conclude the effect of prenatal alcohol exposure on maxillary and mandibular arc measurements is primarily on growth parameters and less directly on maxillary and mandibular growth.

2808W
Genetic analysis of Gonadal Disorders of Sex Development (46,XY DSD) by cytogenetic and molecular methods. A. Shojaei1, F. baghban-aran2, R. Ebrahizmzadeh-Vesali1, F. Behjat3, J. Tavakkoly-Bazza1, 1) Medical Genetics, Tehran University of Medical Science, Tehran, Iran; 2) Genetics Dept, Varamin-Pishva Branch, Islamic Azad University, Varamin, Iran; 3) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Among Disorders of sex development (DSD), 46,XY gonadal dysgenesis is the most complicated, heterogeneous and rare disease. Various genes have been associated with gonadal dysgenesis. Among these, SRY, NRS5A1, DHH, DAX1 and WNT4 genes are performed. In continue MLPA was used to detect deletions and duplications in DAX1 and WNT4 and subsequent imbalances were confirmed by real time PCR. Additionally, other potential loci were investigated by whole genome Array CGH method. In this study, one new chromosomal rearrangement and SRY deletions were found in one and five patients, respectively. Previously described NRS5A1 and DHH allelic variants were observed. A heterozygous partial deletion was presented in NRS5A1 gene and heterozygous partial duplication was found in WNT4 gene. These deletions/duplications were subsequently confirmed by real time PCR. Array CGH results confirmed the chromosomal abnormality in the patient with abnormal chromosome 13, 15 and 18 and showed the exact region of rearrangements. Also, one partial deletion was detected in the SOX2OT gene. Autosomal chromosome abnormalities could also play a role in disorders of sex developments. SRY gene deletion still remains as an important diagnosis of these disorders and has a similar incidence in our patients compared with other reports and should be the first gene for testing in GD. Del/dup mutations found to be more common than point mutations in our patients. Therefore, it might be preferred to check Del/dup mutations prior to check SRY deletion. Further, in cases of gonadal dysgenesis and it should be taken into account in molecular approaches to study GD patients. Array CGH is a valuable tool for finding responsible genes in GD and could unravel some potential loci in this regard.
Beckwith-Wiedemann Syndrome (BWS)(OMIM 130650) is a genetic disorder with complex inheritance mode. Possible patterns include autosomal dominant inheritance with variable expressivity, contiguous gene duplication on chromosome 11p15, and genomic imprinting resulting from a defect or absence of the maternal copy gene. BWS has a prevalence of 1 in 13,700 births; it is characterized by overgrowth, macrosomia, macroglossia, organomegaly, exomphalos and predisposition to embryonal tumor development, most common are Wilms tumor or nephroblastoma. Clinical diagnosis is primarily, requiring 3 major criteria or 2 major and 3 minor criteria. OBJECTIVE: To present a familial case of BWS. CASE REPORT: Family data: Mother with antecedent of 4 pregnancies, 2nd correspond to case 1 and 4th to case 2, she presents ear pits. Case 1: 8 years 5 months female, product of the 2nd pregnancy term, obtained by cesarian section from non-consanguineous parents, with 28 years (he) and 29 years (she) at birth time. Apgar 9-10, weight and height ~pc97, macroglossia, depressed nasal bridge, posterior helical ear pits and exomphalos surgically corrected, presented neonatal hypoglycemia. During the first year of age presented overgrowth, renal ultrasonography reported right pyelic duplication. At the age of 3 years was performed transversal and anteroposterior reduction glossectomy. Actually weight and height p97 and hemihypertrophy was detected. Case 2: Two years 11 months female, product of the 4th pregnancy, obtained by cesarian section at 30 weeks, weight at birth 2500g. Physical examination: weight and height >pct 97, posterior helical ear pits, macroglossia, umbilical hernia; has not required surgical procedure. Renal ultrasonography without CT (computed tomography). Conclusion: Clinical findings of both cases have BWS criteria diagnosis. The treatment consisted of surgical reduction of exomphalos and anterior-transversal reduction glossectomy in case 1. The patient of case 2 has not required surgical intervention. The renal ultrasonography in patient 1 reported right pyelic duplication and in patient 2 renal ultrasonography was normal. The early diagnosis of BWS allows a complete treatment and a genetic counseling. The early and continuous monitoring of glycemia level in the first days after birth provide an opportunity treatment of hypoglycemia and could prevent neurologic complications. The abdominal and renal ultrasonography support early diagnosis of kidney tumor.
2811W
The Coordination of Rare Diseases at Sanford (CoRDS) patient registry for all rare diseases and those undiagnosed. E.A. Donohue, N.A. Simpson, R.M. Bourscheid, D.A. Pearce. Sanford Children's Heath Research Center, Sanford Research, Sioux Falls, SD.

BACKGROUND Sanford Children's Health Research Center at Sanford Research has established a rare disease registry named CoRDS (Coordination for Rare Diseases at Sanford). CoRDS houses de-identified contact and clinical information for patients who have been diagnosed with any rare disease. CoRDS' mission is to accelerate rare disease research by creating a resource of rare disease data for researchers and a mechanism by which participants can be contacted about research opportunities. METHODS CoRDS serves as a rare disease patient registry with several components. First, there is a data collection component, in which the patient-reported data is collected. Second, data management involves the archiving, collation, and accumulation of data. Third, dissemination of data ensures that researchers can access and utilize information for their research. A Scientific Advisory Board has been established to provide oversight and review applications to ensure that researchers have IRB approval and the research proposal is sound. CoRDS utilizes Velos eResearch, a clinical research information system supporting the collection of data and processes in study design/monitoring/execution, patient recruitment, reporting, data integration, compliance and safety monitoring for its software solution. Participant recruitment is closely tied to the process by which CoRDS partners with Patient Advocacy Groups (PAGs) that support patients diagnosed with rare diseases. PAGs communicate enrollment information to their membership and educate their researchers about the opportunity to utilize CoRDS as a resource. RESULTS CoRDS has 853 participants (446 enrolled, 407 in the screening), representing 127 rare diseases. CoRDS has partnered with 44 PAGs and organizations in the medical and rare disease field. These partnerships help increase participant enrollment and disease numbers. DISCUSSION Collecting and collating data on patients diagnosed with any rare disease offers the opportunity to perform a comparative analysis to better understand and treat the diseases. Many treatments are symptomatic, thus treatment strategies for one disease may be beneficial in application to other diseases with similar clinical profiles. CoRDS serves as a source of data to help researchers better understand the diseases. CoRDS is studying and can help them identify potential participants for research studies or clinical trials with the potential to accelerate their timelines.

2812T
Homoplasy of a mitochondrial 3697G>A mutation causes Leigh syndrome. Y. Negishi1, A. Hatton1, E. Takeshita2, C. Sakai2, N. Ando1, T. Ito1, Y. Goto2, S. Saitoh1. 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan.

Leigh syndrome (LS) is a subacute necrotizing encephalomyelopathy characterized by bilateral symmetrical necrotic lesions in the basal ganglia and brainstem. We report that three siblings with LS harbor the m.3697G>A mutation in a homoplasmic fashion. Three siblings from healthy parents are all affected with LS of different severity. Patient 1 is a 9-year-old girl presenting with progressive gait disturbance from 18 months. Currently, she can walk without support, and speak sentences. Patient 2 is a 7-year-old boy. Progressive rigidity of four limbs started at 2 years. Currently, he cannot stand without support and does not have meaningful words. Patient 3 is a 5-year-old girl, who presented with progressive gait disturbance from 18 months. Her condition severely deteriorated after viral infection at 2 years, and has been bedridden with frachestoscopy since then. Brain MRI showed necrotic lesions in the bilateral basal ganglia in all patients, but brainstem involvement only in patient 3. Entire mitochondrial DNA sequence of a biceps brachii muscle specimen from patient 1 revealed that this patient harbored the m.3697G>A substitution (G131S) of the ND1 gene. This mutation was homoplasmic in all three siblings and heteroplasmic in their mother. Heteroplasmic m.3697G>A has been reported in patients with Leber's hereditary optic neuropathy (LHON) and patients with typical deletions of the mitochondrial genome. Histopathology, congenital fiber type disproportion, myotubular/centronuclear myopathy, and central core/multi-minicore disease. The diagnoses are based on histopathological analysis, by which the muscle is reviewed for specific features associated with a congenital myopathy diagnosis, such as central nuclei or nemaline rods. When none of these defining features are present and a distinct congenital myopathy diagnosis cannot be determined based on histopathological analysis, the patient may be said to have an undefined congenital myopathy. We report the retrospective review of 61 patients enrolled in the Beggs Congenital Myopathy Research Program with a diagnosis of undefined congenital myopathy who had pathology reports and muscle biopsies available. After this review, 34 cases were reclassified with a pathological diagnosis of a specific congenital myopathy including centronuclear myopathy, congenital fiber type disproportion, and multi-mini-core disease or another neuromuscular condition like congenital muscular dystrophy. Of the remaining 27 patients, 7 patients had varying clinical presentations including myasthenic syndrome, chronic pain, and facial weakness, whereas 20 patients presented with weakness or hypotonia at early age as a unifying feature. Seven of this latter group had additional findings not related to muscular weakness including developmental delay, autism spectrum disorder, and learning delay. The remaining 13 patients had a clinical history consistent with a congenital myopathy, with delayed motor milestones, feeding difficulties, and diffuse hypotonia, but with normal or nonspecific findings on biopsy. We explore the clinical phenotype associated with the m.3697G>A homoplasmic mutation that is present in the patients originally classified as undefined congenital myopathy.

2814W
Molecular and clinical study of 30 Angelman syndrome patients with UBE3A mutations. K. Hosok1, S. Saitoh1, 1) Department of Pediatrics, Hokkaido University Graduate School of Medicine, Sapporo, Japan; 2) Department of Pediatrics and Neonatology, Nagoya City University School of Medical Sciences, Nagoya, Japan.

Angelman syndrome (AS) is a neurodevelopmental disorder related to genomic imprinting at 15q11-q13. Genetically, AS is classified in five genetic groups; approximately 70% are caused by deletion at 15q11-q13, 5% paternal uniparental disomy 15, 5% imprinting defects, 10% UBE3A mutations, and 10% unknown etiologies. Deletion positive patients have been reported to show more severe clinical features than those with patients in other genetic groups including UBE3A mutation group. In order to delineate molecular and clinical characteristics in AS patients with UBE3A mutations, we reviewed identified mutations in the UBE3A gene in our series of molecular testing for patients with AS, and tried find genotype-phenotype correlation. We detected 26 UBE3A mutations in 30 patients with AS, including 7 familial and 23 sporadic patients. It is of note that 7/26 (27%) mutations were present in exon 16. Frameshift mutations were most predominant (18/26, 69%). Seventy three % of mutations have not been reported previously. Next, we compared clinical features of the patients with UBE3A mutations to those of 10 AS patients with typical 5Mb deletions. There were no significant difference between two groups for major clinical features of AS. However, motor developmental milestones including head control, independent sitting and independent walking were significantly less delayed in patients with UBE3A mutations than in patients with typical deletions. Therefore, our study indicates that patients with UBE3A mutations show milder phenotype compared to patients with typical deletions. Our study also demonstrates three critical genes (GABRB3, GABRA5 and GABRG3) encoding subunits of GABA(A) receptor. Haploinsufficiency of these genes may explain more severe phenotype in patients with typical deletions.

2813F
Defining the Undefined Congenital Myopathies: Pathological Findings and Clinical Features. L.C. Swanson1, P.D.S.C. Claxton1,2, A.H. Beggs1,2. 1) Boston Children's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA.

The congenital myopathies are a group of rare and genetically heterogeneous neuromuscular disorders that affect skeletal muscle, often causing numerous symptoms including, but not limited to, non-progressive muscle weakness, hypotonia, respiratory insufficiency, scoliosis, facial weakness, and feeding difficulties. The severity of symptoms and age of onset are variable among individuals with the same diagnosis. Mutations in different genes can lead to similar histopathological features, and mutations in the same gene can give rise to variable clinical and pathological phenotypes. Several specific congenital myopathy diagnoses include nemaline myopathy, congenital fiber type disproportion, myotubular/centronuclear myopathy, and central core/multi-minicore disease. The diagnoses are based on histopathological analysis, by which the muscle is reviewed for specific features associated with a congenital myopathy diagnosis, such as central nuclei or nemaline rods. When none of these defining features are present and a distinct congenital myopathy diagnosis cannot be determined based on histopathological analysis, the patient may be said to have an undefined congenital myopathy. We report the retrospective review of 61 patients enrolled in the Beggs Congenital Myopathy Research Program with a diagnosis of undefined congenital myopathy who had pathology reports and muscle biopsies available. After this review, 34 cases were reclassified with a pathological diagnosis of a specific congenital myopathy including centronuclear myopathy, congenital fiber type disproportion, and multi-mini-core disease or another neuromuscular condition like congenital muscular dystrophy. Of the remaining 27 patients, 7 patients had varying clinical presentations including myasthenic syndrome, chronic pain, and facial weakness, whereas 20 patients presented with weakness or hypotonia at early age as a unifying feature. Seven of this latter group had additional findings not related to muscular weakness including developmental delay, autism spectrum disorder, and learning delay. The remaining 13 patients had a clinical history consistent with a congenital myopathy, with delayed motor milestones, feeding difficulties, and diffuse hypotonia, but with normal or nonspecific findings on biopsy. We explore the clinical phenotype associated with the m.3697G>A homoplasmic mutation that is present in the patients originally classified as undefined congenital myopathy.
Stargardt Disease, a Clinical Description. L. Mora*, M. Tamayo1,2, F. Rodriguez*, M. Valencia. 1) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá, Colombia; 2) Fundación Oftalmológica Nacional, Bogotá, Colombia.

INTRODUCTION. Stargardt’s disease or juvenile macular degeneration, first described in 1909 by Karl Stargardt, is known as a nosologic entity of autosomal recessive (AR) inheritance. It is the most common cause of macular degeneration in childhood, due to accumulation of lipofuscin in the retinal pigment endothelium and limited energy exchange at level of the photoreceptors; clinically evident between 7 and 12, is shown as a progressive deterioration of visual acuity. It has been reported worldwide incidence of 1/10000, but no data are available at the Colombian population. MATERIALS AND METHODS. * 10 patients (4 men - 6 women) with clinical diagnosis of Stargardt disease. They were obtained from the database of patients with genetics eye diseases of the National Ophthalmological Foundation (FUNDONAL). * Average age now 39.2 years, range 12-69 years. * The patients were evaluated with a complete medical history, fundoscopy under pupil dilation and best corrected visual acuity using Snell scale and fluorescein angiography. * 6 patients were evaluated with OCT (Optical Coherence Tomography) images, which were performed using 5 lines radiated scanner, manually centered on the fovea. RESULTS. * 100% of patients present with best corrected visual acuity of > 20/100. * 80% of those tested were found in stage I / II. * Inverse correlation was found between visual acuity and foveal thickness. * Symptoms start in the first two decades of life in 90% of the patients. * We consider the AR inheritance mechanism as the most likely in the study population. DISCUSSION. Stargardt’s disease is a progressive disease, affecting individuals in the first decade of life. In the clinical and paraclinical description of 10 individuals in collaboration of the Institute of Human Genetics at the Javeriana University and National Ophthalmological Foundation, it was established that for our population the most frequently reported symptoms in the first decade of life include nictalopia and tunnel vision. Symptoms have slow progression, most patients were in stage I and II, this being compatible with a visual acuity of > 20/100, with an average of evolution of 20 years. The OCT findings show foveomacular reduced thickness and loss of photoreceptors in the macular area, which can be useful in the diagnosis of Stargardt disease, it was found strong correlation between foveal thickness and visual function in patients sample.

2816F

Mixed phenotype of incontinentia pigmenti and anhidrotic ectodermal dysplasia with immunodeficiency in a patient with duplication of the IKBKG gene. A. Ramalingam, E. van Asbeck, T.J. Chen, E. Morava. Human Genetics, Tulane University School of Medicine, New Orleans, LA.

We report on a 30 year old female patient with a mixed phenotype of incontinentia pigmenti and anhidrotic ectodermal dysplasia with immunodeficiency in combination with normal intelligence and a progressive neurologic disease. The patient presented with recurrent benign tumors, multiple hemangiomas and cysts and arteriovenous malformation. She has cysts in her ear canal and neurinoma. Besides fatigue, muscle pain and generalized muscle weakness she also developed rapidly progressive peripheral neuropathy, macrocephaly with absence of seizures and hearing loss. She has experienced sudden visual loss on her right eye due to retinal hemorrhage. She presented with progressive problems with gastric emptying and gastro-paresis necessitating a G-tube. Skin symptoms with color changes in a linear pattern, decreased hair growth, patchy alopecia and decreased sweating appeared after puberty. She also has severe and recurrent life threatening infections with compromised immunity but without obvious T or B cell deficiency or low immunoglobulin levels. A duplication of approximately 136 kb at Xq28 was identified by array CGH, and confirmed by qPCR. G6PD and IKBKG genes were fully duplicated while FAM3A and GAB3 were partially duplicated. Deletions, partial duplications and point mutations in IKBKG have been detected in patients with X-linked anhidrotic ectodermal dysplasia with immunodeficiency (XL-EDA-ID). The clinical phenotype of our patient supported our postulation that the fully duplicated IKBKG also played a pathogenic role in XL-EDA-ID patients. Our patient shows a progressive, mixed phenotype of both incontinentia pigmenti and anhidrotic ectodermal dysplasia with immunodeficiency, suggesting more complicated pathogenic mechanisms, such as reported somatic mosaicism, variable expression in different tissues, and skewed X-inactivation. IKBKG is likely to involve in these mechanisms. Further characterization of this duplication and its pathogenic role is in progress.
2817W
Single Molecule Targeted Sequencing of Long Fragments (>1kb) for Ovarian Hypersensitivity Syndrome. F. Orkunoglu-Suer1,2, A. Harratson1, D. Frankfurter1, P. Gindoff1, E. Hoffman1,2, T.J. O’Brien1, 1) Children’s National Medical, Research Institute, Genetics in Medicine, NW, Washington DC, 20010; 2) Dept of Integrated System Biology, Sciences Department, the George Washington University Medical Center, Washington, DC 20037; 3) Department of Pharmaceutical Sciences, Toward J Dunn School of Pharmacy, Shenandoah University, Ashburn, VA; 4) Department of Obstetrics and Gynecology, The George Washington University Medical Center, Washington,.; 5) Department of Pharmacology and Physiology, the George Washington University Medical Center, Washington, DC 20037.

Targeted resequencing using next-generation sequencing technology is being rapidly applied to the molecular diagnosis of genetic diseases. To identify novel predictive genetic biomarkers for ovarian hyper stimulation syndrome (OHSS) that is multi-genic in nature and requires rapid genomic techniques for the identification of risk alleles we employed targeted single molecule sequencing technology. To perform a proof-of-concept study, we selected four patients with OHSS. We developed a custom Targeted Single Molecule Sequencing (T-SMS) Panel; total 44 loci and 3756 primer pairs using long 1 kb standard fragment size with droplet PCR technology (Rain DancerTM). T-SMS was carried out using single molecule real time (SMRT) DNA sequencing (Pacific BiosciencesTM). Secondary analyses were conducted using the Genome Analysis Toolkit for SNP discovery embedded in Smart Portal followed with ANNOVAR for functional annotations. Filtered functional variants further validated using conventional Sanger sequencing in original samples. Target enrichment PCR yielded amplicons averaging 1 kb fragment size. The total targeted sequence was 3.18 Mb per sample. Mean mapped CCS accuracy was 97%, with filtering set default >10X CCS. Using this approach, we have identified a non-synonymous variant LHCGR (NP_000224.2:p.Asn291Ser; rs12470652) in two severe OHSS cases. Critically, combining larger fragments and droplet PCR targeted approach helped us to overcome gDNA and other fragments that may occupy unnecessary sequencing cells, and cause reduction of reads. Single Molecule Sequencing technology provided a new resolution by providing significantly longer reads, reads of GC rich sequences, sequencing of many target genes in a shorter time (minutes instead of hours), low error profile that has promises for clinical diagnosis. Best of our knowledge this is the first clinical study reporting effective sequencing of 1 kb size amplicons utilizing droplet PCR technology. This data show excellent utility for follow up validation and can be further modified and extended to assisted reproductive therapy and infertility genetics.

2818T
Myocardial thickness and ventricular Tei index by echocardiography among normal developing fetuses. Y.-H. Zhang1,2, Q.-Y. Cao1,2, J. Ge1,2, N. Zhong1,2,3,4, 1) Ultrasound, The 4th Municipal Hospital of Shijiazhuang, Shijiazhuang, China; 2) Center of Translational Medicine for Maternal and Children Health, The Maternal Health and Obstetric Hospital, Shijiazhuang, China; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, New York.

Congenital heart disease (CHD) has become the top list of birth defect in China in the recent years. In order to accurately monitor the fetal heart development, normal developmental parameters are necessary to be established. For which, we have undertook a study to obtain the normal fetal myocardial thickness and ventricular Tei index, and to explore its growth regularities as well as reference ranges of normal fetal growth among 257 normal singleton fetuses. These fetuses were randomly selected from normal pregnancies that underwent prenatal ultrasound screening for birth defect in our hospital. The fetal myocardial thickness and ventricular Tei index varied in different pregnant weeks. Fetal myocardial thickness grew with the increase of pregnant weeks and there was a linear correlation between them. Right ventricular wall was thicker than other parts. The ventricular Tei index decreased gradually with the increase of gestational age and Tei index of right ventricle was lower than that of the left ventricle. Fetal myocardial thickness and Tei index could be estimated accurately by echocardiography. Clinical fetal fetuses of different gestational ages are helpful in early detection and timely management of fetal abnormalities.

2819W
Harlequin Ichthyosis: A rare case of two consecutive harlequin ichthyosis in the offspring of a sequindrigavida. O. Franciska, R. Santos. Obstetric Gynecology Department, Jose R Reyes Memorial Medical Center, Rizal Avenue, Manila, Philippines, MD.

ABSTRACT: Harlequin fetus is rare and the most severe form of congenital ichthyosis, inherited as autosomal recessive trait with an incidence of about 1 in 300,000 births. It is characterized by hyperkeratosis and desquamation of the epidermis which begins prenatally. The skin barrier is severely compromised, leading to excessive water loss, electrolyte imbalance, hyperthermia, and increased risk of life threatening infection. Here in our institution, we report a case of a pregnant woman with a previously born child affected with Harlequin Ichthyosis. For this second pregnancy, they were expecting to have a normal baby since the congenital sonographic scan done prenatally was unremarkable. However, the mother delivered at 29-30 weeks age of gestation with the fetus showing signs of Harlequin Ichthyosis. The child was a 1200 gram female neonate born prematurely by partial breech extraction. Clinical manifestation of harlequin ichthyosis were present at birth. Furthermore, the fetus suffered from respiratory distress and was very ill, few hours after delivery and died few days after birth. Prenatal Diagnosis of HI is made with the use ultrasound guided fetal skin biopsy, imaging technique with 2D and 3D real time sonography as well as genetic mutational analysis. Prenatal genetic counseling is very essential in this case because of the serious implications to consider for her offspring. The complex management of our patient can be best achieved by using a multidisciplinary approach characterized by strong communication, both among the medical team and the family. Key words: Harlequin ichthyosis, congenital, prenatal diagnosis, ultrasound.

2820T
Prenatal presentation of fetal anemia associated with e(3.8)-thalassemia: two new cases. E. Gagne1, Y. Sab2, M.F. Delisle1, N. Kent1, T. Nelson1, L.A. Brown1, K. Schliade-Bartusiak1, S. Rogers1, S. Pichard1, L.D. Wadsworth1, N. Au1, 1) Children and Women’s Health Centre of British Columbia, Vancouver, BC, Canada; 2) Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada; 3) Department of Pathology and Laboratory Medicine, University of British Columbia, BC, Canada; 4) Department of Paediatrics, University of British Columbia, Vancouver, BC, Canada.

e(3.8)-thalassemia is rare hematological disease caused by either a deletion of the beta-globin gene cluster, or deletion of the upstream beta-thalassemia (Hb at 9 months 96g/L) and the patient did not require further blood transfusion. Here, we report the clinical course and molecular characterization of two new cases. Case 1- A 26 year girl G1P0, Filipino (Southeast Asian) who presented at 19 weeks and 4 days of gestation with polyhydramnios, microcephaly. Born by Caesarean section at 36 weeks of gestation. Fetal karyotype was normal. Cordocentesis showed a fetal hemoglobin of 30g/L, which required two intra-uterine transfusions. Delivery was induced at 36 weeks of gestation. Cord blood hemoglobin was 150g/L. Hemoglobin levels improved slowly over the 1st 2 months 96g/L and the patient did not require further blood transfusion. Alpha and Beta thalassemia molecular testing was normal. A microarray analysis (Affymetrix T7500 SCAN) revealed a deletion of 54 kb encompassing the LCR, HBE, HBG1 and HBG2 genes. The deletion was confirmed by FISH. The child was shown to be maternally inherited. Case 2- A 42 years girl G2T1L1. Caucasian (Irish/Scottish) who presented at 33 weeks and 3 days of gestation with polyhydramnios, umbilical cord cyst, thick placenta (7.8cm) and elevated MCA peak systolic velocity (94cm/sec). Patient was delivered by cesarean section at 33 weeks and 5 days gestation because of fetal hydrops. Cord blood hemoglobin was 49g/L, and multiple blood transfusions were needed over a 6 months period. Microarray analysis showed a deletion of 771 kb of the entire beta-globin gene cluster. The deletion was shown to be inherited from the father who had no prior history of significant hemoglobin family history. He never revealed a paternal aunt previously diagnosed with e(3.8)-thalassemia 24 years ago and was reported as the >185 kb Canadian deletion (Diaz-Chico, 1988). In conclusion, we report the first e(3.8)-thalassemia deletion diagnosed in a patient of South Asian ethnicity and is the 3rd individual with e(3.8)-thalassemia in the clinical presentation between carriers of the same deletion. In addition, we show that microarray can be a valuable tool to diagnose these deletions.
Aicardi-Goutières syndrome carrier screening in Ashkenazi Jewish population.

2824T Non Invasive Prenatal Screening - Are we providing a false sense of security? S. Klugman, B. Suskin, K. Erskine, N. Kirshenbaum, S. Dolan. 1) Reproductive Genetics, Montefiore Med Ctr, Bronx, NY; 2) Materia

Background: Historically, a patient who has had a high risk for aneuploidy after first trimester screening was offered invasive prenatal testing via chorionic villous sampling or amniocentesis. In the past year, non-invasive prenatal screening has been offered by some in lieu of diagnostic testing. Currently, four companies offer this testing. Case: A 37 year old woman had a first trimester screen at an outside clinic with an increased risk for DS of 1/5 and an increased risk for trisomy 13/18 of 1/25. She was counseled by her primary obstetrics provider and chose to do non invasive prenatal screening. The results were low risk for trisomies 21, 18 and 13. The patient reports she was relieved at that time and chose to continue the pregnancy. The patient was referred to our facility at 19 weeks gestational age for amniocent

Discussion: This case highlights the limitations of non-invasive prenatal screening. It also reiterates the importance of pre-test counseling and post-test counseling when using non invasive prenatal screening as outlined in the recent ACOG/SMFM and ACMG guidelines. Non-invasive prenatal screening tests for a limited number of chromosome aneuploidies, and they differ by platform. In this case specifically, choosing a test that included sex chromosome aneuploidy would have been preferred. In addition, this case highlights the fact that in those patients with a positive first trimester screen who undergo invasive testing, the resultant karyotype is not always consistent with the abnormal screening result. A full karyotype and/or microarray, i.e. a diagnostic test, is recommended after a positive screen and will provide more information than a non invasive prenatal screen.

2825W Two unrelated cases of female fragile X carriers with proximal duplication of non-repetitive DNA sequence within the CGG repeat region of the FMR1 gene. J. Skeen, J. Coppeninger, J. McCarver, R. Cao, S. Filipovic-Sadic, L. Schneitzler, D. Sevilla, J. Weissberger, A. Hardi, S. Nol

Nearly all individuals with fragile X syndrome have a full expansion mutation of more than 200 CGG trinucleotide repeats in the untranslated region of the FMR1 gene. Females with fragile X premutations (55-200 CGG repeats) have a risk of transmitting a fragile X full mutation to their offspring. New high resolution PCR methodologies enable precise sizing of the CGG repeat region and reveal non-repeat sequences, such as AGG interruptions. Here, we report two fragile X carrier females with unexpected DNA sequence findings in the CGG repeat region of the FMR1 gene detected by these novel PCR assays.

Two clinical diagnostic laboratories utilizing the same FMR1 PCR technology independently identified an unusual electrophoretic reading suggesting a non-CGG DNA sequence within the CGG repeat region. These findings were confirmed by DNA sequencing of non-CGG DNA in the affected CGG repeat region. Two unrelated, pregnant fragile X carriers. The samples were further analyzed for AGG interruptions and methylation status, and sequenced across the repeat region. Patient #1 is an African American female who underwent fragile X carrier testing, and was reported to have a FMR1 allele size consistent with 78 repeats, but with an unusual electrophoretic trace suggesting a segment of non-CGG DNA within the CGG repeat region. Patient #2 is a female of Taiwanese descent identified by fragile X carrier testing to have an FMR1 allele size consistent with 58 repeats, but also having evidence of non-CGG DNA present in the repeat region. Sanger sequencing on both maternal DNA samples confirmed the insertion of a unique non-repetitive sequence of 71 nucleotides in Patient #1, and 76 nucleotides in Patient #2. In both cases, the inserted sequence was homologous to a 36 bp sequence within the FMR1 gene that is directly upstream of the CGG repeat region. This type of intra-repeat insertion-duplication has not been previously reported in the medical literature to our knowledge. Both patients have reportedly normal health and intelligence, and both are currently pregnant with male fetuses, one of whom has been confirmed to also carry the typical FMR1 allele without any expansion. These findings suggest that the increasing use of more sensitive, high resolution PCR technologies will reveal rare genotypes with uncertain clinical consequences, raising new questions and challenges for researchers, clinical providers, and patients.

2826T Non-invasive prenatal testing (NIPT): Proceed with caution. A case of trisomy 18 mosaicism in a phenotypically normal newborn. H. Welsh, H. Ardinger, M. Begleiter, L. Zhang, L. Cooley. 1) Department of Pediatrics, Children’s Mercy Hospitals and Clinics, Kansas City, MO; 2) Department of Pathology and Laboratory Medicine, Children’s Mercy Hospitals and Clinics, Kansas City, MO; 3) University of Missouri-Kansas City School of Medicine, Kansas City, MO

NIPT is used to identify fetuses with chromosomal anomalies so parents and their physicians are able to make critical care decisions. The newest screening methods use massively parallel sequencing of circulating fetal DNA to detect chromosome aneuploidy. This non-invasive prenatal screening (NIPT) allows for the detection of chromosome anomalies without the risk of miscarriage. However, as a new front line screening test, a positive NIPT result must be confirmed by conventional methods prior to making major or irreversible medical decisions. We present a 41-year-old woman, who had MaternT21™ testing at 11 weeks gestation due to advanced maternal age. The test was positive for Trisomy 18. Amniocentesis was performed at 15 weeks gestation. FISH analysis of uncultured amniocytes showed three chromosome 18 centromere signals in 2.5% of nuclei. Chromosome analysis found an extra small marker chromosome in one of 15 colonies. FISH analysis of cultured amniocytes found three chromosome 18 centromere signals in 3 of 21 colonies. Serial sonograms were performed with normal fetal anatomy and growth documenting the female child, born at term, weighed 7 pounds 9 ounces, was 19 inches in length, and was non-dysmorphic. Random chromosome analysis of cord blood, cord, membranes and villi found a normal diploid karyotype in 98%, 65%, 100% and 100%, respectively. Two abnormal cell lines were cultured from the extra chromosome 18 cell line, one with complete trisomy 18 (positive for chromosome 18 centromere and 18q21 probes). The extra small marker was present in 2% of blood, and 0% of cord, membranes and villus metaphase cells. Trisomy 18 was absent in 96% of blood, 95% of cord, 85% of membranes and 82% of villus nuclei. This case shows NIPT may have low sensitivity for chromosomal aneuploidy and NIPT should be used as a screening test. Positive NIPT results must be carefully followed up with gold standard prenatal classic cytogenetic diagnostic methods before critical decisions are made. This new methodology may reveal fetuses with mosaic conditions that previously may have gone undetected.


Traditional carrier screening assays can identify only the most common mutations within a gene, typically those recommended by the American College of Medical Genetics, the American Congress of Obstetricians and Gynecologists, and the National Society of Genetic Counselors. Next generation sequencing (NGS) technologies are capable of detecting novel mutations which are only detectable with comprehensive sequencing. NGS has significant advantages over traditional carrier screening and is expected to yield higher detection rates, irrespective of patient ethnicity, resulting in fewer missed carriers.
2828T
1) Medical Genetics Division, CHU Sainte-Justine Montreal, QC, Canada; 2) CHU Sainte-Justine Research Center, Montreal, QC, Canada; 3) Department of Pediatrics, University of Montreal, Montreal, QC, Canada; 4) Dept of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada; 5) Eeyou Awaash Foundation, Chisasibi, QC, Canada; 6) Cree Board of Health and Social Services of James Bay, Chisasibi, QC, Canada.

STATEMENT OF PURPOSE: Cree encephalitis (CE) and Cree leukoencephalopathy (CLE), two severe neurodegenerative autosomal recessive diseases, are found in the James Bay Cree communities (Northern Quebec, Canada). A population-based carrier screening program for CE/CLE (CSP) was developed by local health authorities in collaboration with a community family support group (Eeyou Awaash Foundation). Two groups are targeted: high school students (≥14 years) and women of reproductive age and their partners, mostly in prenatal settings. The CSP has not been evaluated since its start in 2006. OBJECTIVE: To describe screening outcomes and participant knowledge and satisfaction.

METHODOLOGY: We obtained ethics approval to use data collected for management purposes by the CSP (12/2006-04/2013). Available variables include demographic information, carrier screening decision and carrier screening status. We determined uptake of screening and prenatal diagnosis, and carrier rates. Data from school-based and prenatal/preconception screening were analyzed separately. RESULTS: 1340 students were offered screening in school-based CSP. Of those, 650 (48.5%) were screened. Complete data were current for 454 screened students. Most had no known family history of CE (98.5%) or CLE (95.8%). 1.2% were screened in the context of an ongoing pregnancy. 22 students were identified as carriers for CE, 48 for CLE, and one for both. For CLE, one potentially at-risk couple was identified: there was no prenatal diagnosis. In the prenatal/preconception setting, 379 individuals were offered screening. Most had no known previous family history of CE (98.7%) or CLE (98.7%). 326 (86.0%) individuals chose carrier screening for CE and/or CLE. 115 (69.3%) women had an ongoing pregnancy. 45 participants were subsequently identified as CLE carriers and 15 as CE carriers, 10 were carriers for both. Of 8 at-risk couples with an ongoing pregnancy, 6 chose prenatal diagnosis. We estimate the population-based carrier rate as 1/11 for CLE and 1/17 for CE, in contrast to incidence-based carrier rates previously calculated at 1/10 and 1/30, respectively.

CONCLUSION: A high uptake of carrier screening in target populations confirms its acceptability in the communities, as well as individual interest in knowing their carrier status. We plan to evaluate the CSP prospectively to assess screening outcomes and participant knowledge and satisfaction.

2829W
Y-chromosome partial deletions and male infertility in Indian sub-continent. S. Rajender1, D. Jaishwal2, K. Kumar3, D.S. Rani4, K. Singh2, R. Dada5, K. Thangaraj6, 1) Endocrinology, Central Drug Research Institute, Lucknow, UP, India; 2) Banaras Hindu University (BHU), Varanasi, India; 3) All India Institute of Medical Sciences (AIIMS), New Delhi, India; 4) Centre for Cellular and Molecular Biology (CSIR), Hyderabad, India.

Complete AZF deletions are a proven risk factor for male infertility; however, partial deletions are now emerging as risk factors. No concrete data on AZF partial deletions in Indian populations are available. We have analyzed 2184 individuals in three cases-control groups from different corners of the Indian sub-continent. Kolkata cohort included 775 fertile and 287 normozoospermic fertile men from Kolkata and adjacent regions. Uttar Pradesh cohort included 547 infertile and 300 fertile control subjects from Lucknow, Varanasi, and adjoining areas. Delhi cohort included 200 cases and 75 control individuals recruited from Delhi and adjacent areas. All subjects were analyzed for complete and partial deletions in the AZF region using STS and SNV markers recommended by the European Academy of Andrology (EAA). This was followed by meta-analyses on gr/gr and b2/b3 deletions. Classical Y-deletions were observed in a few individuals only. Microdeletions involving the entire AZF region (b2/b4 recombination) were observed in all cohorts and their presence increased the risk of infertility. Most interesting observation concerned gr/gr deletions, which were observed in all cohorts at comparable frequency and increased the risk of infertility. B1/b3 partial deletions were less common and did not affect infertility risk significantly. Interestingly, b2/b3 deletions were absent in most of our cohorts and were not a risk factor for infertility. In conclusion, partial deletions in the AZF region are risk factors for male infertility in Indian populations. The data from these cohorts suggest that gr/gr partial deletions are a strong risk factor and that b2/b3 deletions, which are very common and increase infertility risk in Chinese and few other populations, are almost completely absent in Indian populations. Population specific differences in the susceptibility to the partial deletions and the relation of the partial deletions to infertility risk are evident from comparison of our data with other populations, particularly Chinese/East Asians. Meta-analysis on partial deletion data suggested that gr/gr deletions are important in infertility and may be considered for inclusion in future infertile candidate panels. Additionally, the AZF B1/b3 deletions on b2/b3 deletions found them to be of least significance for testing in a clinical setting.

2830T
MTHFR C677T POLYMORPHISM IN MEXICAN PATIENTS WITH POLYCYSTIC OVARY SYNDROME. J.M. Salazar-Dávalos1, N. Suárez-Magaña2, J.P. Mena3, M. Salazar-Páramo4, E. Chávez4, M.A. Aceves-Aceves1, N.O. Dávalos1, M.G. González-Mercado5, F. Grover-Paz6, I.P. Dávalos1. 1) Instituto de Genética Humana, DGH, Facultad de Medicina, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) UMAE, HGO, IMSS, Guadalajara, Mexico; 3) Doctorado Genética Humana, CUCS, Universidad de Guadalajara; CIBO-IMSS, Guadalajara, Mexico; 4) Depto. Fisiología, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 5) Doctorado Genética Humana, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 6) Depto. Fisiología, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico.

Introduction: Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in reproductive age patients and one cause of anovulatory infertility. There are findings of high homocysteine levels in PCOS, the C677T polymorphism of the MTHFR gene have been associated with hyperhomocysteinemia and decrease of folic acid. Objective: To determine the association between MTHFR C677T polymorphism in Mexican patients with polycystic ovary syndrome. Material and methods: We included 64 patients with PCOS diagnosis according to Rotterdam criteria and 101 Mexican mestizos as reference group (M). The genotyping was performed by PCR/RFLP technique. The M group presented Hardy-Weinberg equilibrium. Results: In the PCOS group (n=64)the MTHFR C677T genotypic frequencies (%(n))were distributed as follows: CC 33 %(21), CT 48 %(31) y TT 19 %(12). The allelic frequency of the C allele was 57 %(73) and T was 43 %(27). The GF in the reference group (n=101) were CC 31 %(31), CT 50 %(51), TT 19 %(19) and AF for C allele 56 %(113) and for T allele 44 %(89). The AF comparison between both groups were not statistically different (p>0.05). Conclusion: The genotypic and allelic distribution between both groups (group-PCOS vs group-M) were similar. This study showed no association between MTHFR C677T polymorphism and PCOS.
Many human disorders result from a complex interaction between an individual’s genetic make-up and environmental stressors. Humans are exposed to numerous xenobiotics constantly and unavoidably such as pesticides, metals, PCBs, etc. Endocrine disruption, genetic predisposition, altered immune surveillance, inflammation and subsequent oxidative stress may antedate adverse reproductive outcomes and contribute to its pathogenesis. Although environmental factors are important, genetics clearly plays a role in adverse reproductive outcomes. Identification of genetic susceptibility variants will lead to better understanding of the role of variable factors in adverse reproductive outcomes. It has been observed that a lot of women with genetic polymorphism do experience normal delivery while some do not. It can be hypothesized that genetic polymorphism requires the presence of certain environmental stimuli to have consequences of clinical significance. The recent abundance of epidemiologic research examining associations between polymorphic genes that code for enzymes involved in xenobiotic biotransformation and disease has on occasion generated interesting findings. Recent studies from our laboratory clearly showed the importance to assess the role of variations in the human genome (polymorphisms) in modifying the effect of exposures to xenobiotics to define ‘Gene-Environment Interaction’, which renders some individuals or groups in the population more or less likely to develop adverse health effects. Current and future efforts to identify new polymorphisms in genes involved in environmental response with larger sample size will broaden the scope of potential genetic effect modifiers. Currently, our laboratory is involved in studying the role of ‘Gene-Environment Interaction’ with reference to xenobiotic metabolism and oxidative stress related genes in various diseases and we have reported the association of xenobiotics with many of adverse reproductive outcomes. Environment Interaction’ with reference to xenobiotic metabolism and oxidative stress. Currently, our laboratory is involved in studying the role of ‘Gene-Environment Interaction’, which renders some individuals or groups in the population more or less likely to develop adverse health effects. Current and future efforts to identify new polymorphisms in genes involved in environmental response with larger sample size will broaden the scope of potential genetic effect modifiers. Currently, our laboratory is involved in studying the role of ‘Gene-Environment Interaction’ with reference to xenobiotic metabolism and oxidative stress related genes in various diseases and we have reported the association of xenobiotics with many of adverse reproductive outcomes. Environment Interaction’ with reference to xenobiotic metabolism and oxidative stress.
The Genetic Predisposition for Uterine Leiomyomata in Recently Admixed Populations: A Preliminary Study in Individuals From Electronic Medical Records. J. Jeff, G. Belfin, D. Rudener-, E. Stah1, S. Purcell-, R. Loos, O. Gottlesman, E. Kenny1,2, 1 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2 Institute for Genomics and Multiscale Biology, Department of Genomic Medicine and Genetics, Icahn School of Medicine at Mount Sinai, New York, NY; 3 Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai; 4 The Mindich Child Health and Development Institute at The Icahn School of Medicine at Mount Sinai.

Uterine leiomyomata (UL) are common benign tumors of the uterus and are the leading cause of hysterectomies. UL affects 25-77% of women and women of African descent have 2-3 fold increased risk. In addition to African ancestry, risk factors for UL include increased age, obesity, endogenous hormonal factors, and family history. Identifying genetic variants in African Americans that predispose to UL may lead to an improved method in future to monitor PIH including preeclampsia using non-peptide biochemical markers.

Identification of differentially expressed non-peptide metabolomic molecules with metabolomics approach in pregnancy-induced hypertensive syndrome. X-C. Tan, X. Hu, J. Pan, X-G. Tao, Y. Lu, Q-X. Shi1,2, N. Zhong1,2,2, 1 Liangyuan Maternal and Children's Hospital, Liangyuan, China; 2 Chinese Alliance of Translational Medicine for Maternal and Children’s Health, Liangyuan, China; 3 Peking University Center of Medical Genetics, Beijing, China; 4 New York State Institute for Basic Research in DD, Staten Island, NY.

Pregnancy-induced hypertension syndrome (PIH) is the most common medical condition encountered during pregnancy and is reported to affect up to 5% of all pregnancies. It is defined as a systolic blood pressure ≥ 140mmHg or a diastolic blood pressure ≥ 90mmHg, and has been shown to account for 24% of all maternal deaths. PIH affects 25-77% of women and is reported to affect up to 5% of all pregnancies. It is associated with increased risk of adverse pregnancy outcomes, including preterm birth, low birth weight, and maternal and fetal morbidity and mortality. Understanding the genetic and molecular mechanisms underlying PIH is important for developing strategies to prevent and treat this condition. Metabolomics is a powerful tool for studying complex diseases such as PIH, and has been used to identify novel biomarkers and therapeutic targets. In this study, we used metabolomics to investigate the metabolic changes associated with PIH, and found that certain metabolites were differentially expressed in PIH patients compared to controls. These findings may provide new insights into the pathophysiology of PIH and could lead to the development of novel diagnostic and therapeutic approaches.

Admixture mapping study of uterine fibroids finds evidence for fibronectin and diabetes genes. K.S. Tsosie1,2, D.R. Velez Edwards1,2, T.L. Edwards1,2, 1 Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2 Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN; 3 Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University, Nashville, TN.

Uterine fibroids are benign tumors that affect up to 77% of women of reproductive age. They are a common cause of infertility and an increased risk of complications during pregnancy, including preterm birth and pre-eclampsia. Understanding the genetic and molecular mechanisms underlying fibroid development is important for developing targeted therapies. In this study, we performed admixture mapping of uterine fibroids using a genome-wide association study (GWAS) approach in African-American (AA) and European-American (EA) women. We identified multiple loci associated with fibroid risk, including one on chromosome 3q26 that is associated with the gene fibronectin. This gene is involved in fibroblast proliferation and has been implicated in the pathogenesis of fibroids. Our findings suggest that genetic factors play a role in the development of uterine fibroids and highlight the importance of understanding the genetic basis of this common condition.
1) Institute of Molecular Medicine & Genetics, Georgia Regents University, Augusta, GA 30912; 2) Division of Basic Medical Sciences, St. George's Medical School, University of London, London SW17 0RE United Kingdom. The hypothalamic-pituitary-gonadal (HPG) axis is regulated by pulsatile secretion of gonadotropin releasing hormone (GnRH). When HPG dysfunction occurs at the level of the hypothalamus, normosmic hypogonadotropic hypogonadism (nHH) results. Patients with IHH manifest absent puberty with low serum levels of sex steroids and gonadotropins. When sense of smell is impaired in IHH patients, it is termed Kallmann Syndrome (KS). Approximately 40% of nHH/KS patients have mutations in a variety of genes that affect GnRH neuron migration and/or GnRH action. One such gene is WDR11, which was identified by positional cloning in a KS patient with a 10;12 chromosome translocation. Approximately 3% of nHH/KS patients had mutations in WDR11, which interferes with WDR11’s binding to EMX1. EMX1 is a transcription factor known to be expressed in the developing forebrain, olfactory sensory neurons, branchial arches, and the embryonic kidney—all of which are affected in KS patients. Recently it was reported that some nHH/KS genes (FGFR1 and NR0B1) could have physiologic effects at both the hypothalamus and gonads. WDR11 is a cytoplasmic protein, which was sequestered in the nucleus following leptinycin B treatment, which inhibits nuclear export. The mechanism of how WDR11 causes KS is not known, but we hypothesize that WDR11 shuttles from the cytoplasm to nucleus, binds with EMX1, and affects expression of downstream genes in GnRH neurons. Therefore, we first wanted to determine if mRNA expression of Wdr11, Emx1, and related gene Emx2 were present in hypothalamic GnRH neurons and gonads. RNA was extracted from mouse GnRH neurons (migratory NLT and postmigratory GT1-7 cells), human GnRH olfactory neuroblasts (FNCB4-hTERT cells), testes, and ovaries, and then subjected to RT-PCR with confirmation by DNA sequencing. In mouse GnRH neuronal cells used. These findings suggest that WDR11 could have functions in GnRH olfactory neuroblasts, Wdr11, Emx1, and Emx2 were expressed. However, in human GnRH olfactory neuroblasts, WDR11 was expressed, but EMX1 and EMX2 were not detected. In mouse testes and ovaries, Wdr11 and Emx1 were expressed at low levels as compared to infertile men. In human GnRH neurons, WDR11 was expressed, but EMX1 and EMX2 were co-expressed in GnRH neurons, testes, and ovaries. Interestingly, EMX1 and EMX2 were not detected in human GnRH olfactory neuroblasts that were derived from an earlier embryological stage than the mouse GnRH olfactory neuroblasts used. These findings suggest that WDR11 could have reproductive functions at both the hypothalamus and gonads similar to other nHH/KS genes.

Understanding the genetics of spermatogenic failure by resequencing the sex chromosomes of infertile men. R. George, J. Hughes, L. Brown, L. Lin, D. Koboldt, R. Fulton, R. Wilson, R. Oates, S. Silber, S. Repping, D. Page, 1) Whitehead Institute, Cambridge, MA; 2) The Genome Institute at Washington University, St. Louis, MO; 3) Boston University School of Medicine, Department of Urology, Boston, MA; 4) Infertility Center of St. Louis, St. Louis, MO; 5) Academic Medical Center, Department of Reproductive Medicine, Amsterdam. The sex chromosomes harbor a large number of genes involved in spermatogenesis and are hemizygous in males. For these reasons, mutations that cause spermatogenic failure—the production of very few or no sperm—are disproportionately likely to be found on the X and Y chromosomes. While several genetic causes of spermatogenic failure, such as microdeletions of the Y chromosome and large-scale cytogenetic abnormalities (e.g. Klinefelter’s syndrome; XXX), have been identified, they only account for 20-30% of cases and the majority of genetic causes remain unknown. To identify new mutations involved in spermatogenic failure, we have captured and sequenced the coding and conserved non-coding regions of the X and Y chromosomes in 300 men with nonobstructive azoospermia and 300 controls. Our targeted regions include 838 and 53 protein coding genes from the X and Y chromosomes respectively, along with 22 Mb of non-coding sequence that contain ncRNAs and putative regulatory regions. We will describe the initial analysis of these sequences, and the identification of putatively causal coding mutations and copy number variants.

Mutations in the Kalikrein related peptidase-3 (KLK3) gene affect semen parameters in Indian men. N. Gupta, D. V. S. Sudhakar, S. N. Sankhwar, N. Gupta, K. Thangaraj, S. Rajender. 1) ENDOCRINOLOGY, CSIR-CDRI, LUCKNOW, UTTAR PRADESH, INDIA; 2) CENTRE FOR CELLULAR AND MOLECULAR BIOLOGY, HYDERABAD, ANDHRA PRADESH, INDIA; 3) DEPARTMENT OF UROLOGY, KING GEORGE MEDICAL UNIVERSITY, LUCKNOW, UTTAR PRADESH, INDIA; 4) INSTITUTE OF REPRODUCTIVE MEDICINE, KOLKATA, WEST BENGAL, INDIA. Kalikrein related peptidase-3 (KLK3) is a highly abundant serine protease in the prostatic epithelium and seminal plasma. It has been known to play a role in fragmentation of semenogelins, resulting in the dissolution of semen coagulum and activation of sperm progressive motility. Among male reproductive organs, KLK3 expression was earlier thought to be restricted only to the prostate gland. Recent studies have reported KLK3 expression in adult testis, suggesting its possible participation in the process of spermatogenesis. Genetically compromised activity of KLK3 might alter spermatogenesis, but analysis of KLK3 sequence to identify mutations in infertile men has not been undertaken. Therefore, we have analyzed the coding region of KLK3 in ethnically matched 875 infertile and 230 fertile men. Interestingly, this study identified three substitutions, of which thirteen were novel (not yet available in public databases). None of these substitutions has been studied in correlation with male infertility. Among the thirty three substitutions, nine resulted in non-synonymous changes while rest of them were either in the intronic regions or resulted in synonymous changes. In silico analysis on novel substitutions revealed four to be missense variants, six to be intronic variants, one to be a 5’UTR variant, one to be a 3’UTR variant, and one to be a downstream gene variant with reference to the transcript ENST00000326003. Variant effect prediction analysis on the non-synonymous substitutions identified p.Met1393Lys, p.Glu174Lys, and p.Ile179Thr to be deleterious. Statistical comparisons of genotype frequencies showed that six SNPs were differentially distributed between fertile and infertile men. The frequency of genotypes ‘GA+AA’ (rs266881, OR = 2.92, P = 0.0001), ‘TC+CC’ (rs11573, P = 0.022), ‘GA+AA’ (rs266875, OR = 1.44, P = 0.016), and ‘TC+CC’ (rs174776, OR = 1.91, P < 0.0001) were significantly higher in infertile men as compared to fertile controls. The genotypes ‘TC+CC’ at intrinsic SNPs c.206+c.235T>C (OR = 0.44, P = 0.002) and ‘GA+AA’ at c.631-74G>A (OR = 0.26, P = 0.02) were more frequent in fertile controls as compared to infertile men. In conclusion, our findings suggest that one or more SNPs in the KLK3 gene may lead to altered semen parameters and affect male fertility; highlighting the importance of this gene in spermatogenesis/testicular function.
Expression Quantitative Trait (eQTL) Mapping in Mid-Secretory Phase Endometrial Cells Identifies Candidate Genes for Recurrent Early Pregnancy Loss (REPL). C.L. Kagan, G. Kosova, K. Patterson, M.D. Stephenson, C. Ober. 1) Dept. of Human Genetics, University of Chicago, Chicago, IL; 2) Dept. of Obstetrics and Gynecology, University of Chicago, Chicago, IL.

Fertility traits are heritable in humans; however, little is known about the genes that influence reproductive outcomes or genetic variants that contribute to inter-individual differences in these traits. To address this gap in knowledge, we performed an unbiased genome-wide study to identify functional (regulatory) SNPs in endometrial biopsies collected during the mid-secretory phase, the time at which implantation occurs, in 34 women of European ancestry with a history of at least two documented intrauterine miscarriages of less than 10 weeks of gestation. Gene expression was measured in RNA using Illumina Human HT12v4 arrays; DNA samples were genotyped using the Affymetrix Axiom™ Genome-Wide CEU 1 Array, which contains 674,517 SNPs. To identify cis eQTLs, we included SNPs within 200kb of each gene and considered an additive model as implemented in Matrix eQTL. Transcripts from 10,191 genes were detected as expressed in these samples. Of 532,538 association tests, 1,088 cis eQTLs in 326 genes were significant at a false discovery rate (FDR) of <5%. We then genotyped the 80 most significant cis eQTLs in 256 women with REPL (not included in the eQTL study) and 232 women with at least one successful pregnancy and no history of infertility or miscarriage (fertile controls), and tested for differences in genotype frequencies between the two groups using the Cochram-Armitage trend. 64 of these SNPs were successfully genotyped using iPLEX (Sequenom). Nine SNPs had p-values < .05 (based on permutation) when only 3.2 were expected by chance, reflecting a significant enrichment for small p-values (p = 0.0037). The most significant association was with a SNP (rs56274787; p = 0.0087) that is an eQTL for the TBCD gene, which encodes a tubulin-folding protein that promotes epithelial cell detachment. Overall, 9 SNPs showed independent evidence both for being functional in the endometrium (i.e., associated with gene expression) and for being significantly associated with REPL patients compared to fertile controls; this study demonstrates that eQTL mapping in a relevant tissue is a useful strategy for identifying genes and associated alleles that are risk factors for REPL, which affects approximately 5% of couples. The eQTLs identified in our study could also serve as candidates for other reproductive disorders such as primary infertility and preeclampsia. Supported in part by NIH grant HD21244.

Clinical Analysis of Relationship between Sperm DNA Damage and Sperm Parameters. F. Kaplan, S. Aydos, B. Altnok Zaim, Y. Yükselen, A. Sungurul, K. Aydos. 1) Ankara University Faculty of Medicine, Department of Medical Biology Ankara, Turkey; 2) Ankara University Faculty of Medicine, Reproductive Health Research Center Ankara, Turkey.

Purpose: Although fertilization rate by using intracytoplasmic sperm injection (ICSI) is increasing, the probability of fertilization with DNA damaged sperm creates risk for both pregnancy success and also health of the new born. Recently, disputes whether sperm parameters give an opinion related to sperm DNA damage or not have become more of an issue. The aim of the study is to investigate the relationship between sperm DNA damage and sperm parameters. Methods: 298 infertile men, between 24-38 years, whose sperm motility percentage and sperm counts are known, were included to the study. 140 of them had also sperm morphology records. In all couples, female factor was eliminated. DNA damage was studied with Comet assay. Observations were made at magnification 400 X using an epifluorescent microscope. 100 cells were analyzed visually from each 3 slide. Each image was classified according to nucleus scale and tail length given a value of 0.1, 2, 3 or 4 (from undamaged class 0 to maximally damaged class 4), so that the total DNA damage score of the slide could be between 0 and 400 arbitrary units (AU). Sperm DNA damage ratios were correlated with sperm motility, sperm count and sperm morphology. Statistical analysis was made with Mann-Whitney u and Kruskal-Wallis tests. Results: Our results showed a significant negative correlation between the sperm counts ≥ 15 million/ml (p<0.001), progressive motility ≥52% (p<0.05), and total sperm DNA damage score. Total sperm DNA damage scores of the samples with sperm count ≥15, progressive motility ≥52%, and sperm DNA damage ≥22.4±5.88 were found 22.4±5.88 respectively. According to our findings there was no correlation between Kruger morphology results and sperm DNA damage scores. Discussion: According to our results, there was no relation between sperm morphology and sperm DNA damage. We found significant relation between sperm count, motility, and total sperm DNA damage score. The significant increase of spermatozoa rates with DNA damage in infertile males comes into prominence due to being some obstacles by using assisted reproductive technology (ART) in these infertile males. Indeed, in case of natural fertilization, spermatozoa reached the fallopian tube have much healthier DNA when compared to which cannot. The results of our study emphasize the risk of using sperm cell with DNA damage in males having abnormal sperm parameters, because of randomly chosen morphologically normal sperms during IVF/ICSI.

Quantitative Analysis of Mixtures by Deep Sequencing of HLA Gene Amplicons Using Next Generation Systems. B.N. Hoglund, M. Rastrou. D. Goodridge, H.A. Erlich, C.L. Holcomb. 1) Human Genetics, Roche Molecular Systems, Pleasanton, CA; 2) Conexio Genomics, Perth, Australia; 3) Children’s Hospital Oakland Research Institute, Oakland, CA. Purpose: The clonal property of next generation sequencing systems allows for the quantitative analysis of mixed samples by simply counting sequences corresponding to the component alleles. Previously, we developed a method using 454 amplicon sequencing and Conexio software for high resolution and high throughput genotyping of the HLA class I and class II loci. This system was used to analyze the blood of a severe combined immunodeficiency (SCID) child and estimate the proportion of maternal cells by counting HLA-C allelic sequence reads corresponding to the non-transmitted maternal allele. Here we report the development of a system that allows us to quantify HLA allelic mixtures in plasma. Methods: DNA from plasma or contrived mixtures of cell lines was amplified using primers that targeted a short region (~150 bp to amplify small DNA fragments in plasma) of HLA DPB1 or DQB1 exon 2. Amplified products were further amplified by emulsion PCR and sequenced on a 454 GS FLX or GS Junior. Sequences were examined using modified Conexio Assign ATF 454 software. This software allows for fast digital analysis of each DPB1 or DQB1 allele. We also analyzed mixtures using the Illumina MiSeq. Results: Using the modified Conexio software, minor HLA alleles in a mixture were readily identified and separated from background ‘noise’ i.e., sequences generated by PCR or sequencing errors. In mixtures of two heterozygous cell lines, the minority HLA alleles could be detected at 0.5% with 1 ng (~140 diploid genomes) DNA input. In one plasma sample from a pregnant woman in the third trimester, fetal HLA alleles were detected at 11.6% of the total of maternal plus fetal alleles. Conclusion: We have developed a system that allows the sensitive and precise analysis of mixtures of HLA. This method could be useful in numerous clinical applications; we have demonstrated the quantification of fetal DNA in maternal plasma during pregnancy by detection of HLA alleles. Such quantification has potential uses in noninvasive prenatal diagnostics.
2843W

Objective: Our primary objective was to establish a cut-off value for the soluble fms-like tyrosine kinase 1 (sFlt-1)/placental growth factor (PIGF) ratio measured using the Elecsys assay to predict late-onset preeclampsia in low-risk women. A secondary objective was to evaluate the ability of combination models that included Elecsys data, second trimester uterine artery (UTA) Doppler, and fetoplacental proteins measurements during Down syndrome screening. Methods: A prospective cohort study was carried of 262 women at low risk for preeclampsia. Maternal plasma samples were obtained for measurement of pregnancy-associated plasma protein-A (PAPP-A), alpha-fetoprotein, unconjugated estriol, human chorionic gonadotrophin, inhibin-A, and the sFlt-1/PIGF ratio. All women underwent UTA Doppler at 20-24 weeks of gestation. Results: Of the 262 women, eight (3.0%) developed late-onset preeclampsia. ROC curves demonstrated that the sFlt-1/PIGF ratio in the third trimester yielded the best detection rate (DR) at a fixed false positive rate (FPR) of 10%. This was followed in rank order by the sFlt-1/PIGF ratio in the second trimester, sFlt-1, and then PIGF. Binary logistic regression was used to determine the five best significant combination models for screening. For a FPR of 5 and 10%, the combination of PAPP-A with the sFlt-1/PIGF ratio in the second trimester yielded a DR of 87.5%, identical to that of serial sFlt-1/PIGF ratios for late PE. The combination of maternal BMI and second trimester sFlt-1 yielded a DR of 87.5% at a 10% FPR. Combining PAPP-A with inhibin-A yielded a 50% DR at a 10% FPR. The combination of PAPP-A and the sFlt-1/PIGF ratio in the third trimester yielded a DR of 62.5% for late preeclampsia at a 10% FPR. Conclusion: The combination of the second trimester sFlt-1/PIGF ratio and PAPP-A data or BMI, and the second trimester sFlt-1 can predict late onset preeclampsia more effectively than any single marker alone.

2844T
First live birth in Hong Kong after preimplantation genetic diagnosis on a disease-predisposition mutation carrier with a novel genomic deletion in BRCA2. Q. Wang,1,2 J. Chowienczyk,2 Y. Tsang,3 W. Yeung4, E. Lau,5 V. Lee,1,2 E. Ng,1,2 P.C. Ho.1 1)Dept. of Obs. & Gyn., Queen Mary Hospital, Hong Kong; 2) Dept. of Obs. & Gyn., The University of Hong Kong, Hong Kong; 3) Dept. of Obs. & Gyn., University of Sao Paulo, Sao Paulo, SP, Brazil; 4) FMVZ - University of Sao Paulo, Sao Paulo, SP, Brazil; 5) FMVA - UNESP, Araçatuba, SP, Brazil

Preimplantation genetic diagnosis (PGD) is an option for couples bearing disease-causing mutations to avoid the bearing of an affected child. The use of PGD on disease-predisposition mutation carriers, such as BRCA mutation is controversial. Here we report the first live birth in Hong Kong, after PGD upon a request from a woman with an early onset of breast cancer. She carries a paternally derived mutant allele with an unknown genomic breakpoint. Due to the limited availability of familial data, sperm haplotyping was conducted on the woman’s carrier brother, leading to the identification of the mutant haplotype. Embryo biopsy was performed on 8 good-quality embryos on day 3, followed by whole genome amplification and linkage analysis with 4 microsatellite markers and 2 intragenic SNP markers. Among the five unaffected embryos identified, one morula and 1 blastocyst were replaced on day 5, resulting in a singleton livebirth. To increase the diagnostic accuracy in future cycles and to better understand the cancer risk among the family, efforts were made to identify the mutation breakpoint, which was found to be novel. The data showed that a genomic sequence of 2596 nucleotides including exon 15 and 16 was deleted. In conclusion, the first live birth in Hong Kong after PGD on disease-predisposition mutation marked the beginning of a new phase of PGD practice in Hong Kong.

2845W
Embryo SNP array genotyping: a model for preimplantation diagnosis in human. F. Campagnini1,2, Y.T. Utsunomiyaz, A.S. Carmo,1,4 J.A. Vinson,2 J.F. García,1 C. Rosenberg1, R.V. Alonso1,4 1) Genetics and Evolutionary Biology Dept, University of Sao Paulo, Sao Paulo, SP, Brazil; 2) FMVZ - University of Sao Paulo, Sao Paulo, SP, Brazil; 3) FMVA - UNESP, Araçatuba, SP, Brazil; 4) Deoxy Biotecnologia, Aracatuba, SP, Brazil

Preimplantation diagnosis and screening is gradually been incorporated to routine for couples with increased risk of particular disorders or to increase in vitro fertilization (IVF) success by selecting euploid embryos. Developments in animal reproduction biotechnology, such as embryo in vitro production (IVP), microinjection and preimplantation genetic diagnosis (PGD) can be used as models for application in human PGD. The aim of this study was to perform PGD in bovine embryos through SNP arrays genotyping (BeadChip BovineLED - 6,909 SNP). The small amount of genomic DNA (gDNA) obtained from embryo biopsy is the main limitation for SNP array analysis. Whole Genome Amplification (WGA) (Repli-g® Mini Kit, Qiagen, Hilden, Germany) was used to increase the amount of gDNA from embryo biopsy and allow the analysis of thousands of SNP simultaneously. Eighty-eight IVF bovine embryos were subjected to microinjection by microaspiration, forming three groups based on numbers of biopsied cells: G1) 05-10 (n=28); G2) 10-20 (n=37); and G3) > 100 - hatched blastocyst (n = 23). All samples were subjected to the same WGA protocol, and 4µL of each sample were genotyped on iScan/Illumina platform. The genotyping quality was assessed using the Call Rate (CR), GenCall Score (GC10), Allele Drop In (ADI) and Allele Drop Out (ADO). Kruskal-Wallis test was applied to investigate differences in the distribution of variables among the groups. Spearman’s rank coefficient was calculated to assess the correlation of each possible variables pair. The results showed a positive correlation between CR and GC10 (0.99/P <0.001), while ADI and ADO rates were negatively correlated with CR and CG10 (ADI/CR: -0.87; ADO/CR: -0.87; ADO/GC10: -0.88; P<0.001 for all variables. Kruskal Wallis pointed to significant differences in all variables (CR, GC10, ADI and ADO) among the 3 groups of biopsies (G1, G2 and G3). The CR average was 59.26 percent; 78.47 percent; and 95.97 percent; for G1, G2 and G3, respectively. Revealing that the number of cells recovered from the embryo biopsy is retaining satisfactory results in SNP analysis. The development of an algorithm based on Mendel’s first law (Law of Segregation) increased CR average to 79.69 percent; 88.20 percent; and 97.28 percent; for groups 1, 2 and 3, respectively. The present study showed that, based on animal results, it is possible to apply SNP arrays in embryo stage, enabling early genotyping for PGD.

2846T
KaryoLite® - A rapid single cell screening assay to simultaneously detect aneuploidies for all chromosomes from whole-genome amplified DNA from 3 day blastomeres. S. Dalilah, R. Walker, M. Scherner. PerkinElmer, Molecular Diagnostics, 940 Winter Street, Waltham, MA, 02451.

Aneuploidy is a genetic defect where the number of chromosomes is either greater or less than 46. Aneuploidies have been detected in a majority of miscarriages. The ability to test for aneuploidies prior to IVF implantation may allow better management and outcome of the pregnancy. The FISH and microarray tests that are used in clinical labs to detect aneuploidy on all chromosomes are impractical for many labs because of cost, labor, and time-to-result. We have developed the KaryoLite® kit, which is a bead-based multiplex assay in PerkinElmer’s BACs on Beads® family, which will be able to detect aneuploidies for chromosomes 1-22, X and Y. KaryoLite® BeOs utilizes a new concept of composite beads having DNA from three different BAC clones on each bead type. The composite clone format expands the region of chromosomal DNA interrogated by each bead. The assay is fluorescence based and uses encoded multiplex beads which have been coupled to BAC derived DNA from defined loci on all 24 chromosomes. This allows detection of aneuploidies on all chromosomes from a clinical sample in a single well of a 96-well microtiter plate with results in as little as 16 hours. A study was performed using 177 DNA samples derived from blastomeres at day 3. DNA from the biopsied cells was processed by whole genome amplification. 240ng of the amplified DNA from each sample was fluorescently labeled and analyzed. The assay was robust with typical signal/background greater than 4, and the KaryoLite assay was completed within 24 hours. Results indicated 88% male and 49% female samples. Of these, 10 (6%) were confirmed normal, 87 indicated whole chromosome aneuploidies, 9 with arm specific aneuploidies and 4 were X and 1 was Y. These results show KaryoLite to be a sensitive, multiplex, high throughput assay that can detect all aneuploidies on all chromosomes from single cell samples. This demonstration uses KaryoLite as a useful tool in preimplantation genetic screening research.
2847W

Quebec perspectives on the medical and social uses of preimplantation genetic diagnosis and on the current service deliveries. F. Duplais- Lefrançois, R. Drouin, C. Bourbarr: Division of Genetics, Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Quebec, Canada.

Since 2010, Quebec’s healthcare system has covered costs related to preimplantation genetic diagnosis (PGD) for the purpose of detecting specific chromosomal or single-gene defects. However, this measure has several shortcomings. In order to identify problems and to better understand the situation, we consulted the Quebec geneticists, obstetricians-gynecologists and genetic counselors likely to be involved in PGD. Among the objectives of this study, we wanted to know their positions on the medical and social uses that could justify PGD, and hear their proposals to improve the current conditions of the delivery of services.

Methodology: Qualitative Research Design – Online questionnaire containing 34 questions has been completed by 15 obstetricians, 15 geneticists and 17 genetic counselors engaged in activities related to prenatal diagnosis in the province of Quebec. Mixed data analysis: quantitative and descriptive for closed questions; qualitative (inductive) and thematic approach for open questions.

Results: The participants propose to: 1) certify laboratories, 2) restrict PGD to lethal or disabling diseases, 3) proscribe PGD for multifactorial diseases or social uses, 4) direct the decision-making process of the patients, 5) create a provincial discussions table or committee to organize the development and the implantation of PGD 6) regulate, standardize and make PGD more accessible. As for conditions that may justify PGD: 1) obstetricians seem less restrictive, 2) geneticists more conservative and 3) genetic counselors more focused on genetic counseling and the needs of the patients.

Conclusion: Despite certain differences between professionals, there is a clear consensus on the necessity of improving the conditions of PGD services, and of reducing problems of accessibility. All also agree on the need to regulate and standardize PGD and to ensure the availability of knowledge relating to the situation.

2848T

Notchless Impacts Multiple Signaling Pathways During Pre-Implantation Development. C.-L. Lo1,2, A.C. Lossie1, 2, J.B. Sherrill1, 2

1) Department of Animal Science, Purdue Univ, West Lafayette, IN; 2) Dept of Medicine, Indiana University School of Medicine, Indianapolis, IN; 3) PULSe Interdisciplinary Graduate Program, Purdue University, West Lafayette, IN; 4) Department of Biological Sciences, Purdue University, West Lafayette, IN.

Per-implantation is a critical stage in mammalian development. Genes and genetic pathways that are necessary for establishing and maintaining maternal-fetal interactions play crucial roles in this process. Several members of the WNT pathway are detected in normal embryos and maternal tissues during this transitional stage. Recent studies in our laboratory demonstrated that mutations in Notchless (Nle1), a member of the vast WD40-repeat family and a putative NOTCH signaling molecule, disrupts expression of several genes in the Wnt pathway. Mutant embryos also express high levels of Cdk9, a downstream target of CDKN1A that promotes cell cycle recovery from replication arrest. At E3.5, mutant embryos showed reduced steady-state levels of Cdk9, suggesting that they are under severe cellular stress that ultimately leads to caspase-mediated apoptosis at the hatched blastocyst (E4.5) stage. To determine if mutants were undergoing cell cycle arrest prior to apoptosis, we examined expression of Cdk9, a downstream target of CDKN1A that promotes cell cycle recovery from replication arrest. At E3.5, mutant embryos showed reduced steady-state levels of Cdk9, suggesting that they were in cell cycle arrest at this stage. Since mutant embryos showed evidence of apoptosis at E4.5 and TRP53 induction often precedes apoptosis, we analyzed Trp53 expression at the mRNA and protein levels. Although we did not detect altered mRNA levels of Trp53 by qRT-PCR, immunofluorescence studies using an antibody that detects an active form of TP53 (Acetyl K376), demonstrates that the acetylated form is only detected in mutant blastocysts at E3.5 and E4.5. Together, these data suggest that at E3.5, Nle1 mutants are undergoing cell-cycle arrest, while at E4.5 the mutants initiate caspase-mediated apoptosis. These two mechanisms are regulated by activation of TRP53. Our data implicate Nle1 in WNT signaling and cell cycle arrest and apoptosis via TRP53-mediated signaling. Intriguingly, WNT signaling is critical for gastrulation in mice. Deletion of Wnt3 leads to failure prior to primitive streak formation, and multiple ligands and receptors are detected in blastocysts and the uterus during peri-implantation.

2849W

The first report of a viable, 35 week gestation pregnancy following the transfer of a genetically normalized blastocyst; embryo normalization can occur during differentiation to the blastocyst stage. W.G. Kearns1, 2, M. St. Amant1, 2, B. Welch3, J. Carter1, A. Potts1, P.R. Brezina1, 2, A.T. Benner1, K.J. Tobler2, G.R. Cutting3, R.P. Dickey4, 1) Center for Preimplantation Genetics, LabCorp, Rockville, MD; 2) Department of Gynecology and Obstetrics, Division of Reproductive Endocrinology and Infertility, Johns Hopkins Medical Institutions, Baltimore, MD; 3) Women’s Hospital, Baton Rouge, LA; 4) The Fertility Institute of New Orleans, Mandeville, LA; 5) Fertility Associates of Memphis, Memphis, TN; 6) McKusick-Nathans Institute of Medical Genetics, Johns Hopkins Medical Institutions, Baltimore, MD.

Case Report. Here we report the progress of a genetically normal 35 week fetus that resulted from the embryo transfer of a genetically normalized blastocyst. A 36 year old female with 9 years of secondary infertility was treated with intravenous fertilization and preimplantation genetic screening (PGS). Two cleavage stage embryos underwent embryo biopsy by laser and a single cell was removed from each embryo for PGS. The single cells underwent lysis and DNA whole genome amplification using a modified random priming method. Following this, comparative genomic hybridization (aCGH) microarray analysis was performed. The single cell from each cleavage stage embryo was aneuploid; embryo one was 48, X, +17, +19, and embryo two was 47, XY, +15. During further embryo culture, only two embryos differentiated to the blastocyst stage of development and underwent a trophectoderm (TE) biopsy and repeat aCGH analysis. The aCGH results showed embryo two to be euploid, 46, XY. Because no cleavage stage euploid embryo was available, the ‘normalized’ embryo was transferred after obtaining informed consent from both parents regarding the risk of delivery of a fetus with Down syndrome. At 15 weeks gestation, diagnosis was determined to be Notchless (Nle1) based on aCGH microarray analysis performed at 16 weeks; confirmed a normal karyotype: 46, XY. Due to the aneuploidy in the single cell from the cleavage stage embryo, uniparental disomy analysis was performed to confirm that one copy of chromosome 21 came from each parent. Masked aCGH microarray analysis of the cleavage stage embryo DNA and the trophectoderm DNA confirmed karyotypes of 47, XY, +15 and 46, XY respectively. The pregnancy is progressing normally at 35 weeks. DNA fingerprinting analysis is ongoing to confirm that the transferred, genetically normalized blastocyst resulted in the viable pregnancy. This report supports the hypothesis that mosaic cleavage stage embryos, with aneuploid cells, can normalize into euploid blastocysts capable of viable pregnancies.

2850T

Stability Testing of a Noninvasive Prenatal Test (NPT) in a Clinical Setting - the MaterniT21™ PLUS Laboratory-Developed Test. R.C. Tim1, J.A. Tynan1, T.J. Jensen1, L. Cagasan3, V. Lu1, L. Liu1, S. Sovath1, M. Rivera1, P. Oeth1, M. Ehrlich1, 1) Sequenom Center for Molecular Medicine, LLC, San Diego, CA; 2) Sequenom, Inc., San Diego, CA.

Excellent clinical performance of a noninvasive test for fetal aneuploidies using next-generation sequencing has been demonstrated by several laboratories. This study demonstrates the reproducibility of genetic processes involved in such an assay, the MaterniT21™ PLUS laboratory-developed test, focusing on the dynamic and stability of the results. The study was divided into two portions to determine the robustness of the MaterniT21™ PLUS test. The first portion of the assay involved preparation of maternal cfDNA isolated from a pool of circulating cell-free (ccf) plasma DNA isolated from women at increased risk for fetal aneuploidy with a known euploid fetus as determined by fetal karyotyping. For these experiments, ccf DNA obtained from 976 women was combined and used to prepare over 1000 libraries. These libraries were sequenced and analyzed for variability of chromosomal representation.

The second portion of the study was designed to investigate the stability of the post-PCR workflow processes. For this part of the study, a total of 1,394 embryos were used for use throughout the entire series of experiments, comprised of 44 known euploid and 44 known trisomy 21 samples. For each experimental subset, all factors were kept constant, including operator, reagent lots, and instruments, except for the particular variable of interest. All flow cells from both portions of the study were clustered to the Illumina® cBot in 12-plex and sequenced on the HiSeq™ 2000 (Illumina, San Diego, California). Sequencing reads were demultiplexed, and aligned to the human genome with Bowtie2 and chromosomal representations were determined. Results demonstrate that while library concentrations and normalized reads per sample were consistent across samples, the stability of the post-PCR workflow processes involved in the assay is remarkably stable for the pooled maternal ccf DNA samples processed by multiple operators, across library batches, and measured on multiple sequencing instruments. From the second part of this study, no significant variability in the workflow process was determined using flow cells as a function of library storage time, flow cell storage time, reagent lot, or sequencing instrument. Both sensitivity and specificity for each of these experimental subsets were determined to be nearly 100%. This study demonstrates the stability of the usage of the MaterniT21™ PLUS test across operators and instruments and reveals the low variability for discrete process steps of the assay.
POSTERS: PRENATAL, PERINATAL AND REPRODUCTIVE GENETICS

2851W Identification of Common Pathogenic CNV by arrayCGH in Prenatal Cases with Oral Clefts. Y. CAO1, Z. LI1, J. ROSENFELD2, A. PATEL3, J. HUANG4, X. SUN5, B. Skotko6, S.W. CHEN7-12. 1) Department of Obstetrics and Gynecology, The Chinese University of Hong Kong, Hong Kong, China; 2) Women and Men's Health, Montefiore Medical Center, Bronx, NY; 3) Department of Obstetrics and Gynecology, The Chinese University of Hong Kong, Hong Kong, China; 4) CBUHK-Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; 5) Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China; 6) Medical Genetics Laboratories, PerkinElmer, Inc., Spokane, Washington, USA; 5) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 6) Medical Genetics Laboratories, Baylor College of Medicine, Houston, TX, USA.

PURPOSE: Oral clefting is one of the most common congenital malformations. With improvements in sonography, it is now being detected more frequently and accurately during routine fetal morphology scan. Prenatal identification of this defect raises questions about prognosis, as oral cleft may be part of syndromic presentations, associated with other congenital malformations, possibly detectable by ultrasound; and/or with neurodevelopmental problems, which could not be assessed prenatally. Identification of a genetic cause may help clarify the prognosis and improve clinical management. A systematic review showed that up to 7.9% of oral cleft cases which were combined with different cleft categories both detected prenatally and postnatally were identified with chromosomal defects by karyotype. However the role of submicroscopic chromosomal variations (CNVs) still remains in limited knowledge in prenatal oral cleft cases. METHOD: We performed a retrospective multi-centre study of prenatal cases undergoing clinical microarray-based comparative genomic hybridization (aCGH) with different categories of oral cleft. RESULTS: There were 119 isolated oral cleft cases and 459 prenatal cases with multiple anomalies. Among these subgroups: isolated cleft lip (CL) (45/278, 16.2%); isolated cleft lip and palate (CLP) (174/278, 26.6%); syndromic CL (46/278, 16.5%); and syndromic CLP (113/278, 40.7%). Cardiac defects, limb defects and brain abnormalities were the most frequent dysmorphic features. The overall detection rate of pathogenic CNVs was 12.9%. Pathogenic CNVs were more commonly identified among syndromic groups, (29/159, 18.2%), especially in syndromic cleft lip, while less (7/119, 5.8%) among isolated cleft. Four pathogenic CNVs were commonly identified across 11 cases: 1q24.1 deletion, 16q12.1 duplication and 11q23.3 deletion recurrent in our cases. CONCLUSION: aCGH is a powerful tool to detect submicroscopic pathogenic CNVs in prenatal cases with oral clefts. It is highly recommended and should be a first-tier test prenatally, especially for suspected syndromic oral cleft cases.

2853W Exploring Placental Gene Expression Pattern in Abnormal Fetal Growth. A. Sabri1,2, C.H.M. Ng3, C. H. M. Ng3, L. D. Lai4, A. D’Silva3, J. Kaur1, J. A. Yeung1, Z. L. Li3, B. Skotko3. 1) Department of Obstetrics, Gynecology and Neonatology, Queen Elizabeth II Research Institute for Mothers and Infants, The University of Sydney, Sydney, Australia; 2) Molecular Biology Facility, Bosch Institute, The University of Sydney, Sydney, Australia; 3) RPA Women and Babies, Royal Prince Alfred Hospital, Sydney, Australia.

OBJECTIVE: Extremes of fetal growth are associated with increased perinatal mortality and morbidity and a higher prevalence of cardiovascular disease, obesity and diabetes in later life. The placenta plays a pivotal role in fetal growth and development and is a potential target for perinatal gene expression studies. We aimed to identify changes in placental gene expression in term pregnancies with evidence of growth dysfunction and to identify candidate genes that may be used to identify abnormal patterns of growth prior to delivery.

MATERIALS AND METHODS: Placenta samples were collected and classified from pregnancies that were small for gestational age (SGA; <10th centile; n=5), large for gestational age (LGA; >90th centile; n=6) or normal (AGA; 40-60th centile; n=5). All pregnancies were 39 weeks’ gestation, with male infants delivered by caesarean section. RNA was extracted from the placental samples prior to microarray gene expression analysis (Affymetrix HG-U219 array). Microarray data were analysed using Partek Express and Ingenuity Pathway Analysis. Significant differential gene expression was defined by >2 fold-change with p<0.05.

RESULTS: Significant differential gene expression was found in both SGA and LGA placentas, with up-regulation in 68 and 18 genes and down-regulation in 9 and 9 genes respectively. Differential expression in SGA placenta was seen in genes involved in embryonic and connective tissue development and differentiation, whereas in LGA placenta genes involved in cardiovascular and haematological disease, cell-mediated immune responses, endocrine and developmental disorders. Individual findings of interest include gremlin 2 (GREM2) and zinc finger protein 711 (ZNF711) in SGA and leptin (LEP) and matrix metallopeptidase 12 (MMP12) in LGA, playing important roles in embryonic and organ development, cell proliferation and tissue differentiation.

CONCLUSIONS: The placental growth differential is associated with differential placental gene expression and affects genes involved with a whole spectrum of developmental and cellular functions. GREM2 and ZNF711 appear to be good candidates for prediction of SGA in term pregnancies. LEP appears to be decreased in LGA infants, which may be of significance for long term outcome.

2852T Assessing the utilization and distribution of an evidence-based resource recommended in the 2013 American College of Medical Genetics and Genomics statement on noninvasive prenatal screening for fetal aneuploidy. S. Meredith1, B. Skokoto2, C. Brasington3. 1) Human Development Institute, University of Kentucky, Lexington, KY; 2) Massachusetts General Hospital Down Syndrome Program, Boston, MA; 3) Down Syndrome Clinic at Carolinas Medical Center, Charlotte, NC.

Introduction and Purpose: The purpose is to share data for the first time on the utilization and distribution of an evidence-based resource as implemented by medical providers delivering a prenatal diagnosis of Down syndrome (DS). A single education resource was prepared with assistance from representatives of ACMG, ACOG, NSGC, and the national DS organizations. It is part of the National Center for Prenatal and Postnatal DS Resources at the University of Kentucky and was released as a recommendation in the 2013 ACMG policy statement. Specifically this assessment determines: 1) Trends in US geographic areas 2. Professional discipline trends 3. Referral sources used by medical providers 4. The impact of professional guidelines, professional newsletters, and internet promotions Methods: This study assessed the number of medical providers who requested/purchased printed copies of “Understanding a DS Diagnosis.” When requesting a free printed book, medical providers are required to submit their name, email, medical practice, address, area of practice, and can optionally indicate how they learned about the book. The data was collected and organized from 681 medical requests/purchases from June 1, 2012-June 1, 2013 to show the number of medical providers from different disciplines requesting books from different regions. Results: The data reveals that the top states where medical providers request the DS educational resource is California (30%), Texas (12%), New York (7%), Massachusetts (6%), Illinois (6%), Pennsylvania (6%), Florida (5%), Wisconsin (5%), and Missouri (5%). The top discipline requesting books is genetic counselors (15%) while the least requests come from OB/GYNs (4%) and Family Practitioners (20%). Data indicates that the majority of requests obtained from professionals via professional organizations, colleagues, and email. Conclusion: The top states requesting booklets typically include higher population states and states where local DS organizations and state health departments are more active in distributing books. Genetic counselors are most likely to utilize a patient resource and receive the most DS training. Yet, OB/GYNs and family practitioners are least likely to utilize the book and receive the least training about DS. Data indicates that ‘word of mouth’ and professional organizations are the most important sources in the medical community for obtaining resources, and online professional newsletters prompt the most requests.

2854T Diagnostic dilemma: fetal cardiomyopathy presenting at a late gestational age. B. Suskin1,2, K. Bajaj1,2, M. Rosner3,2, P. Dar1,2, S. Klugman1,2. 1) Department of Obstetrics & Gynecology and Women’s Health, Montefiore Medical Center, Bronx, NY; 2) Albert Einstein College of Medicine, Bronx, NY.

Background: Cardiomyopathies are diseases of the myocardium that affect the mechanical or electrical function of the heart. American Heart Association has stated that most are due to genetic causes. Elucidating a cause, expediting testing and explaining a prognosis is challenging in the fetal setting. Case: A 30 year-old multiparous woman at 30 weeks with gestational diabetes was referred from an outside clinic for fetal surveillance. Her pregnancy was otherwise uncomplicated and she denied any significant family history. Oligohydramnios was found. At her follow-up ultrasound with fetal medicine specialists, oligohydramnios was seen again. They also found a grossly enlarged heart, a small pericardial effusion and an increased PR interval and suspected cardiomyopathy. The fetus did not have hydrops. Fetal echocardiograms by pediatric cardiologists confirmed these findings. TORCH and other in the mechanical or electrical function of the heart. American Heart Association has stated that most are due to genetic causes. Elucidating a cause, expediting testing and explaining a prognosis is challenging in the fetal setting. Case: A 30 year-old multiparous woman at 30 weeks with gestational diabetes was referred from an outside clinic for fetal surveillance. Her pregnancy was otherwise uncomplicated and she denied any significant family history. Oligohydramnios was found. At her follow-up ultrasound with fetal medicine specialists, oligohydramnios was seen again. They also found a grossly enlarged heart, a small pericardial effusion and an increased PR interval and suspected cardiomyopathy. The fetus did not have hydrops. Fetal echocardiograms by pediatric cardiologists confirmed these findings. TORCH and other potential causes were ruled out. The infant was lost to follow-up, but represented in labor. A female infant was delivered at 39 weeks at 3585 grams with apgars 5/6/8 at 1, 5 and 10 minutes respectively. She was intubated and transferred to the pediatric hospital. Echocardiograms show bilateral ventricular hypertrophy and moderately depressed function. The work-up is pending. Discussion: Due to the fact that the patient presented at 36 weeks with sparse prenatal care, the work-up was limited to infectious and maternal causes. Neonatal cardiomyopathies can be seen in high-output failure, anesthesia, volume overload and myocardial damage. A common non-genetic cause of fetal cardiac disease is maternal hyperthyroidism secondary to maternal hyperthyroidism. There are also many possible genetic causes. These include syndromic, metabolic, neuromuscular, mitochondrial and sarcomere mutations. Despite our abilities to evaluate for many such causes, we are often limited prenatally to the setting by time and cost. It is important to consider both genetic and acquired causes of cardiomyopathy of fetuses during pregnancy. The increased risk of fetal demise is known, especially when hydrops is present. This poster presents the genetic considerations, work-up and limitations of prenatally diagnosed fetal cardiomyopathy.


2855W

Improving efficiency and cost of next generation sequencing of maternal cell free DNA for the detection of fetal aneuploidy: L. Chitty1,2, J. Weir1, C. Boustred1, S. Fielding1, F. McKay1, H. White2, Z. Kingaby1, S. Humphreys2, E. Tsoaga2, N. Lench3, J. Betley2.

1) Clinical Molecular Genetics Unit, UCL Institute of Child Health, London, London, United Kingdom; 2) Illumina Inc, Chesterford Research Park, Little Chesterford, Essex, CB8 1XL, UK; 3) North East Thames Regional Genetics Laboratory, Great Ormond Street NHS Foundation Hospital, London, London; 4) Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, Wilts, SP2 8BJ.

We have recently reported preliminary data describing a new PCR-free method of preparing cDNA for next generation sequencing (NGS), which may remove PCR-bias and potentially improve accuracy of trisomy 13 and 18 detection as well as decreasing time and costs for NGS of maternal plasma cfDNA.

As decreasing time and costs for NGS of maternal plasma cfDNA.

2856T

Cordocentesis: an alternative prenatal procedure for women who missed amniocentesis in developing regions of China. Q. Cao1,2, J. Gu1,2, Y.-Y. Peng1,2, E.C. Jenkins1, W.T. Brown1, N. Zhong1,2,3,4.

1) The 4th Municipal Hospital of Shijiazhuang, Shijiazhuang, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children’s Health, China; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in DD, Staten island, NY.

Objective: To assess the efficacy of cordocentesis for rapid karyotyping in a high-risk obstetric population. Methods: Cordocenteses were performed on 64 pregnant women with different indications for prenatal diagnosis in an outpatient setting from November 2009 to September 2012. Fetal chromosome some karyotypes were examined. Results: The most frequent indication for cordocentesis was a fetal abnormality on sonography (92.2%). All of the procedures were performed using the free cord loop. The success rate of cordocentesis was 92.2% (59/64) and cell culture was 94.9% (56/59). Transient bleeding was common at the puncture site, but did not stop spontaneously within one minute. Seven chromosomal abnormalities (12.5%) and 4 chromosomal variations (7.1%) were identified among 56 specimens cultured. Six of the seven chromosomal abnormalities (11.8%) were identified among 50 specimens with normal ultrasound findings. Trisomy 21 and prevalent chromosomal abnormality, was present in normal pregnancies when available. Several fetuses’ DNAs (16/23) were analysed by microarray. Out of the 23 cases, 15 are inherited, 6 are de novo and 2 of unknown origin. Termination of pregnancy was mainly due to ultrasound findings and occurred in 7 cases. We discussed the potential role of cordocentesis in this syndrome and about different hypothesis to explain its clinical heterogeneity. We examined in particular, the co-existence of additional CNVs and their contribution to the phenotypic variation in the 22q11.2 duplication syndrome. Conclusion: We report the first prenatal microarray analysis of patients with 22q11.2 duplication diagnosed prenatally would help the understanding of this pathology.

2857W

Single nucleotide polymorphism (SNP)-based non-invasive prenatal testing (NIPT) detects triploidy: two case studies. Z. Demko, M.P. Hall, M. Hill, B. Zimmermann, S. Sigurjonsson, M. Rabionowitz, Natra Inc., San Carlos, CA.

Objective: To report two NIPT cases with evidence of triploidy. Design: Two triploid samples were analyzed using the non-invasive prenatal Next-generation Aneuploidy Test with Single Nucleotide Polymorphisms (NIPT-SNPs) assay from a commercial protocol (reporting detection of Trisomies 13, 18, and 21, and Monosomy X) or as part of an IRB-approved research protocol. Materials and Methods: Cell-free DNA (cfDNA) isolated from plasma from pregnant women was amplified via a modified version of the TruSeq PCR targeting 19,488 SNPs. Sequencing results were analyzed with the NATUS algorithm. Laboratory personnel (and the NATUS algorithm) were blinded to sample karyotype. Results: NATUS analysis identified more than two fetal haplotypes present for multiple chromosomes, indicating twins or triploidy. Case 1 (fetal fraction: 6.4%): multiple paternal haplotypes on multiple chromosomes (13, 18, 21, X) were detected, indicating either paternally-inherited triploidy or fraternal twins. Case 2 (fetal fraction: 20.8%): multiple paternal haplotypes on multiple chromosomes (13, 18, 21, X) were detected, similarly indicating either a paternally-inherited triploidy or fraternal twins. Ultrasound with chorionic villus sampling (Case 1) or amniocentesis (Case 2) confirmed single gestations affected with triploidy. Conclusions: This SNP-based approach facilitates the detection of additional alleles at polymorphic loci, thus not requiring a reference chromosome, and therefore opens the potential ability to detect triploidy. Here, this method correctly flagged both triploidy cases. This method is capable of distinguishing twins from triploidy based on the presence (in the case of twins) or absence (i.e., triploidy) of two fetal haplotypes. Since the pathophysiology of triploidy is particularly clinically relevant as it is correlated with partial molar pregnancy, and molar pregnancy (partial or complete) can develop into hydatidiform mole. Studies are ongoing to differentiate all twin and triploidy cases; in the interim ultrasound will be required to confirm which cases may be at risk of triploidy. Support: NIH R44HD062114-02.
2859W
Validation of epigenetic marker for noninvasive prenatal diagnosis of fetal trisomy 18. DE. Lee1, SY. Kim1, JH. Lim1, HJ. Kim1, SY. Park1, HM. Ryu11, 2 1) Laboratory of Medical Genetics, Cheil General Hospital and Women’s Healthcare Center, Seoul, Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women’s Healthcare Center, Kwanbong University College of Medicine, Seoul, Korea.

Background: The quantification of cell-free fetal DNA by methylation-based DNA discrimination has been used in noninvasive prenatal diagnosis of fetal chromosomal aneuploidy. Maspin (SERPINB5) gene, located on chromosome 18q21.33, is hypomethylated in the placenta and completely methylated in maternal blood cells. The objective of this study was to evaluate the accuracy of noninvasive fetal trisomy 18 detection using tissue specific-methylation of maspin in the first trimester of pregnancy. Methods: A nested case-control study was conducted using maternal plasma samples collected from 66 pregnant women carrying 11 trisomy 18 and 55 normal fetuses. Using real-time quantitative methylation-specific PCR, the concentrations of unmethylated maspin (U-maspin) and methylated maspin (M-maspin) were measured in first trimester maternal plasma. Results: U-maspin concentrations were significantly elevated in women with trisomy 18 fetuses compared with controls (96.9 vs 19.5 copies/mL, P<0.001). The specificities of U-maspin and M-maspin concentration for noninvasive fetal trisomy 18 detection were 96.4% and 74.5%, respectively, with a sensitivity of 90.9%. In the risk assessment for fetal trisomy 18, the adjusted odds ratios of U-maspin and M-maspin concentration were 325.2 (95% confidence interval: 17.9-5903.8, P<0.001) and 19.0 (95% confidence interval: 2.1-175.6, P=0.009), respectively. Our results suggest that U-maspin and M-maspin concentration may be useful as potential biomarkers for noninvasive fetal trisomy 18 detection in the first trimester of pregnancy, irrespective of the sex and genetic variations of the fetus.

2860T
Highly accurate non-invasive detection of fetal aneuploidy for chromosomes 13, 18, 21, X and Y. B. Levy1, S. McAdoo2, B. Zimmermann3, M. Banjevic4, B. Pettersen5, H. Hall6, Z. Dekom7, M. Hill8, M. Rabinowitz1. 1) Department of Cell Biology and Pathology, Columbia University, New York, NY; 2) Natera, Inc, 201 Industrial Rd, San Carlos, CA, 94070.

Objective: To develop a non-invasive prenatal test (NIPT) that will detect fetal chromosome 13, 18, 21, X, and Y copy number through analysis of cell-free (cfDNA) isolated from maternal plasma using the Next-generation Aneuploidy Testing Using SNPs (NATUS) algorithm. Methods: cfDNA was isolated from 699 maternal plasma samples at ≥9 weeks gestation, including 70 aneuploid samples, under an institutional review board approved protocol. DNA at 19,488 polymorphic loci covering the target chromosomes was amplified in a single multiplex-PCR assay and sequenced. Results were analyzed using the NATUS algorithm, which uses Bayesian statistics to calculate sample-specific accuracies and detects chromosomally abnormal fetuses from maternal plasma with high accuracy at chromosomes 13, 18, 21, X and Y as early as 9 weeks gestation.

2861W
Disease-specific characteristics of fetal epigenetic markers for non-invasive prenatal diagnosis of trisomy 21. J. Lim1, D. Lee1, S. Park1, D. Kim1, H. Kim1, H. Ahn1, S. Lee1, K. Choi1, M. Kim1, H. Ryu1,2 1) Laboratory of Medical Genetics, Medical Research Institute, Cheil General Hospital and Women’s Healthcare Center, Seoul, South Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women’s Healthcare Center, Kwanbong University College of Medicine, Seoul, South Korea.

BACKGROUND: Non-invasive prenatal diagnosis of trisomy 21 (T21) is being actively investigated using fetal-specific epigenetic markers (EPs) that are present in maternal plasma. Recently, 12 EPs on chromosome 21 were identified based on tissue-specific epigenetic characteristics between placenta and blood, and demonstrated excellent clinical performance in the non-invasive detection of fetal T21. However, the disease-specific epigenetic characteristics of the EPs have not been established. Therefore, we validated the disease-specific epigenetic characteristics of these EPs for use in non-invasive detection of fetal T21. METHODS: We performed a high-resolution tiling array analysis of human chromosome 21 using a methyl-CpG binding domain-based protein (MBD) method with whole blood samples from non-pregnant normal women, whole blood samples from pregnant normal women, placenta samples of normal fetuses, and placenta samples of T21 fetuses. Tiling array results were validated by bisulfite direct sequencing and qPCR. RESULTS: Among 12 EPs, only four EPs were confirmed to be hypermethylated in normal placenta and hypomethylated in blood. One of these four showed a severe discrepancy in the methylation patterns of T21 placenta samples, and another was located within a region of copy number variations. Thus, two EPs were confirmed to be potential fetal-specific markers based on their disease-specific epigenetic characteristics. The MBD results of these EPs were consistent with the results obtained by bisulfite direct sequencing and qPCR. However, the methylation level of these EPs showed inter-individual variability. CONCLUSION: We validated that two EPs have the potential to be fetal-specific EPs which is consistent with their disease-specific epigenetic characteristics. The findings of this study suggest that disease-specific epigenetic characteristics should be considered in the development of fetal-specific EPs for non-invasive prenatal diagnosis of T21.

2862T
The concerns of health professionals and pregnant women involving the non-invasive prenatal diagnosis of trisomy 21 in Quebec and in France. A.K. MAGLO1, P. DROUIN2, J.M. MOULTOUN1, C. BOUFFARD4 1) Division of Genetics, Department of Pediatrics, Université de Sherbrooke, Faculty of Medicine and Health Sciences, Sherbrooke, Quebec, Canada; 2) Department of Obstetrics and Gynecology, Université de Sherbrooke, Faculty of Medicine and Health Sciences, Sherbrooke, Quebec, Canada.

Introduction: In France, the Agence Nationale de Sécurité du Médicament currently assesses the relevance of authorizing the non-invasive prenatal diagnosis (NIDP) through blood testing of trisomy 21 (T21) (PrenaTest®). As Quebec begins to be confronted with this issue, our objective was to identify and compare the concerns of health practitioners, pregnant women and couples involving the NIDP of T21 from probative data available for France and Quebec. Methodology: Qualitative aspect: 1) Systematic narrative review: research through key-words in probative data (scientific publications and gray literature) in databanks: PubMed, ERIC, JSTOR, SCOPUS, etc. 2) Thematic and comparative analysis of the content: coding, categorization and classification of information in themes, allowing for the identification of the variables and relationships established in the texts. 3) Comprehensive synthesis. Results: In Quebec, reflection focuses mostly on the invasive diagnosis of T21. In France, the debate extends to issues related to NIDP. In both countries, health practitioners, pregnant women and their partners are equally concerned with 1) the lack of information on NIDP and 2) the importance attributed to genetic counseling. Finally, health practitioners have questions relating to 3) standards of practices and 4) ethics; pregnant women have questions relating to 5) decision-making autonomy and 6) the attitude of society towards individuals living with T21. Conclusion: Parallel to the evolution of NIDP techniques, we continue to lack probative data that would allow us to establish health policies ensuring that this change in paradigm in the field of prenatal diagnosis does not diminish the quality of services and decision-making autonomy of women, and does not encourage discrimination against people with T21.
2863W
A single nucleotide polymorphism-based approach to non-invasive prenatal testing identifies lingering cell-free fetal DNA in pregnancies with vanishing twins.


Objective: Using single nucleotide polymorphism (SNP)-based non-invasive prenatal testing (NIPT) to analyze cell-free fetal DNA (cfdNA) allows for identification of samples with multiple embryos and/or vanishing twins; these additional parental haplotypes distinguish single from multiple gestations. Vanishing twins may confound fetal copy number calling in singleton pregnancies due to cfdNA from the vanishing twin placenta. A SNP-based approach allows detection of vanishing twins in reportedly singleton pregnancies, and may reduce the frequency of incorrect results. We present three such cases. Methods: Commercial samples were processed at a single reference laboratory. Isolated cfdNA was amplified using multiplex PCR targeting 19,488 SNPs covering chromosomes 13, 18, 21, X, and Y. Sequencing data was analyzed using Next-generation Aneuploidy Test Using SNPs (NATUS) algorithm that uses Bayesian statistics to analyze multiple copy number hypotheses and determine the Maximum Likelihood hypothesis, calculating chromosome-specific accuracies without requiring a reference chromosome. Follow-up clinical information was sought for cases in which NATUS was suggestive of a vanishing twin. Results: Case 1: vanishing twin reported at 6w0d, maternal blood drawn at 1iw3d; Case 2: twin loss reported at 6w5d, maternal blood drawn at 13w5d. Neither case had evidence of a second sac at blood draw. Case 3: two sacs but one fetus reported at 12w6d, maternal blood drawn at 19w1d. Conclusions: This method detected cfdNA from a vanishing twin as many as 8 weeks after demise. The contribution of cfdNA from a vanishing twin increases the cfdDNA estimate, potentially resulting in analysis of samples with cfdNA levels from the surviving twin below that which is appropriate for accurate detection of aneuploidy. Additionally, residual cfdDNA from a vanishing twin could be aneuploid (explaining the cause for the loss) whereas the surviving twin may not be aneuploid. Together, this may result in incorrect conclusions.

2865W
Experience using a rapid assay for aneuploidy and microdeletion detection in over 2900 prenatal specimens.


While microarray testing can identify chromosomal abnormalities missed by karyotyping, its prenatal use is often avoided in low-risk pregnancies due to the possible identification of variants of uncertain significance (VOUS). We tested 2971 samples using a rapid, BACs-on-Beads™-based assay with probes for sex chromosomes, common autosomal aneuploides, and a panel of 15 microdeletion syndromes, and 5 reciprocal microduplication syndromes, designed as an alternative to microarray testing in low-risk pregnancies and an alternative to rapid aneuploidy FISH in high-risk pregnancies that also undergo microarray analysis. Interpretable results were obtained in 2942 cases (99.0%), with 89% receiving results the next business day after sample receipt. Aneuploides were detected in 7.3%, with the rate dependent on indication for study (p<0.0001). Partial chromosome abnormalities were detected in 0.5% (n=14), including six low-risk cases referred for maternal age, abnormal maternal serum screen, or isolated soft ultrasound markers. One VOUS was obtained (0.3%), which was determined to be benign following further characterization by microarray. Abnormalities were confirmed through secondary testing; neither false negatives nor false positives were found, within limitations of the test. Female polyplody cannot be detected, while polyplodies with Y chromosomes are suspected and confirmed through additional analysis. This assay allows for detection of clinically significant microdeletions and aneuploidy with rapid results. It tests for more conditions than rapid aneuploidy FISH, though cannot detect all polyploidy. When combined with karyotyping, this assay provides increased interrogation of specific chromosomal regions, while limiting the identification of VOUS.

2865T
Methods for fetal fraction quantification in circulating cell-free DNA sequencing libraries.

J. Tyman1, G. Hogg1, J. Fox1, P. Iyer1, M. Ehrich2, 1) Sequonem Center for Molecular Medicine, San Diego, CA; 2) Sequenom, Inc., San Diego, CA.

OBJECTIVE: Circulating cell-free (ccf) fetal DNA fraction in maternal plasma can be measured using chromosome Y markers from a male fetus, fetal specific methylation markers, or paternally inherited alleles absent in the maternal genome. Here, we compare the quantification of ccf fetal DNA fraction in maternal plasma sequencing libraries using single nucleotide polymorphism (SNP) allele frequencies, chromosome Y representation by sequencing, and the ratio of chromosome Y and total genomic copies by droplet digital PCR (ddPCR). METHOD: Maternal plasma DNA was isolated and plasma DNA libraries were prepared using TruSeq reagent kits (Illumina, San Diego, California), clustered at 12-sample plexing and sequenced on the HiSeq 2000 (Illumina) for 36 cycles. Plasma DNA sequencing reads were aligned to the human genome (hg19) and the proportion of reads derived from chromosome Y were used to estimate fetal fraction. Maternal plasma DNA libraries were also used as template for a multiplexed single-tube PCR SNP panel and a duplex ddPCR assay. The multiplexed single-tube PCR assay amplified 67 high minor allele frequency SNPs and was designed such that a subsequent universal PCR incorporated indexed adapter sequences to allow multiplexed sequencing on the MiSeq (Illumina). The duplex ddPCR assay was designed to quantify SRY copies arising from the male fetal genome and TERT total genomic copies on the QX100 (Bio-Rad, Hercules, California). RESULTS: By sequencing 84 plasma DNA libraries for chromosome Y representation, 38 samples passing assay QC metrics were derived from a donor carrying a male fetus. Sequencing based chromosome Y fetal fraction significantly correlated to ddPCR based fetal fraction (r2=0.72), but ccf fetal fraction showed significant correlation to SNP based fetal fraction (r2=0.63). Sequencing based chromosome Y fetal fraction showed significant correlation to SNP based fetal fractions (r2 = 0.81). CONCLUSION: These data demonstrate concordance of ccf DNA fetal fraction measurement in sequencing libraries by multiple methods. SNP allele frequencies, chromosome Y representation showed higher correlation. The ddPCR method likely is negatively impacted by sampling error of SRY copies leading to a higher variability of fetal fraction quantification. Of course, only the SNP allele frequency assay is able to quantify fetal fraction in pregnancies carrying a female fetus.
Prenatal MLPA screening and aCGH analysis detected cytogenomic abnormalities in four cases with fetal ultrasound anomalies. J. Xie1, Z. Xu2, Q. Geng2, F. Xu4, P. Li1. 1) Center for Prenatal Diagnosis, Shenzhen Maternity and Child Healthcare Hospital, Shenzhen, Guangdong, China; 2) Department of Genetics, Yale School of Medicine, New Haven, CT.

Ultrasound detected fetal anomalies have been the most important clinical indications for prenatal genetic diagnosis. However, the associations of specific ultrasound findings with most syndromic cytogenetic disorders remain elusive. We have applied rapid MLPA screening and aCGH analysis to high risk pregnancies with abnormal ultrasound findings. Cytogenomic abnormalities were characterized in four cases. The first fetus with ultrasound findings of growth retardation, bilateral cleft lip and palate, right-sided aortic arch and multiple cystic hygromas was detected with a 46,701 Mb duplication of 8q22.3-q24.3 and a 23,839 Mb deletion of 7q33-q36.3. Follow up parental study found the father to be a carrier of a balanced 7q33q8q22.3 translocation. The second fetus with single-ventricle heart, transposition of the great arteries and double superior vena cava was diagnosed as Jacobsen syndrome (OMIM#147791) by a de novo 13 Mb distal deletion of 11q24.1-q25. The third fetus with split-hand-split-foot malformation and multiple umbilical cord cysts was diagnosed as type 1 split-hand-foot malformation (SHFM1, OMIM#183600) with a de novo 19,971 Mb interstitial deletion of 7q11.23-q21.3. The fourth fetus with partial hydrops, a hypoplastic cerebellum and a large ventricle showed a 699.8 Kb deletion including the TERT gene at 5p15.33 for the diagnosis of Cri du Chat syndrome (OMIM#123450). Post-test genetic counseling was performed with detailed genomic information and well characterized postnatal syndromic features. All parents made an informed decision to elect termination. Prenatal ultrasound findings reported in literature for Jacobsen syndrome, SHFM and Cri du Chat syndrome were reviewed and summarized. Our results demonstrated that comprehensive evaluation of abnormal prenatal ultrasound findings and their association with known cytogenomic disorders will lead to better practice of prenatal genetic testing and counseling.

Chinese Alliance of Translational Medicine for Maternal and Children’s Health (CATMMACH): A unique resource for longitudinal cohort study of pregnancies. J. Pan1,2, Q.-X. Shi1,2, Y. Gu1,2, X.-C. Lou1,2, M.-F. Hua1,2, N. Zhong1,2,3,4,5, 1) Lianyungang Maternal and Children’s Hospital, Lianyungang, Jiangsu, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children’s Health (CATMMACH), China; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in DD, Staten Island, NY. Globally, 536,000 women died from causes related to pregnancy and childbirth in 2005, and 3.7 million newborn infants died in 2004. Globally, an estimated 15 million infants are born preterm annually, and more than one million infants die annually from preterm birth complications, making these complications the most frequent cause of all deaths worldwide of children younger than 5 years of age. Recent translational research to elucidate the molecular mechanism underlying preterm births is focusing on two major categories: conventional research to explore biomarkers, which may be applied to early prevention of preterm birth, and an integrated systems biology approach with “omic” technology, to uncover the genetic risk factors, including gene mutation, genetic predisposition or environmental-genetic interaction. However, the molecular mechanisms underlying preterm birth are not yet clearly understood because it is a complex disorder. Many environmental and genetic factors are involved, and gene-gene and gene-environmental interaction(s) likely play key roles. To study these roles, a national network in China, the Chinese Alliance of Translational Medicine for Maternal and Children’s Health (CATMMACH), has been developed, comprising eight hospitals and two medical genetics centers. Collection of quantitative specimens from pregnancies, along with detailed clinical information, has become a critical resource to enable the biobanking program of this alliance to conduct a longitudinal cohort study of pregnancy outcomes. Availability and accessibility of the pregnancy specimens from the biobank of the alliance allow international and domestic investigators to perform translational research to uncover the molecular pathophysiology of pregnancy and its outcomes.

Integrative transcriptome analysis reveals dysregulation of canonical cancer molecular pathways in placenta leading to preeclampsia. R. Moslehi1, J.L. Mills2, C. Signore3, A. Kumar1, X. Ambroggio4, A. Dzutsev5, 1) Epidemiology and Biostatistics, Cancer Genomics Center, University at Albany, State University of New York, NY; 2) Epidemiology Branch, Eunice Kennedy Shriver National Institutes of Child Health and Human Development, NIH, Bethesda, MD; 3) Pregnancy and Perinatology Branch, Eunice Kennedy Shriver National Institutes of Child Health and Human Development, NIH, Bethesda, MD; 4) Bioinformatics and Computational Biosciences Branch, National Institutes of Allergy and Infectious Disease, NIH, Bethesda, MD; 5) Cancer Inflammation Program, Center for Cancer Research, National Cancer Institute, NIH, Frederick, MD.

Background Based on our clinical observations and subsequent genetic epidemiologic studies of DNA repair disorders, we previously identified associations between specific nucleotide excision repair (NER)/transcription gene mutations in the fetal genome and the risk of placental maldevelopment and preeclampsia, possibly due to impairment of Transcription Factor (TF)IIH-mediated functions in placenta. To identify the underlying mechanisms, we designed an integrative analysis of relevant transcriptome data sources containing gene expression arrays of fetal cell-derived tissue. This novel approach has enabled enrichment of causative mechanisms which underlie the previously-reported associations. Methods We conducted meta-analysis of gene expression patterns in placenta from four case-control studies of preeclampsia. A preeclampsia-specific gene list obtained from this meta-analysis was then interrogated in three relevant data sources (normal and time course placentas (i.e., placentas from first, second and third trimester pregnancies), hypoxic trophoblasts and XPD fibroblasts (i.e., cells predisposed to preeclampsia)). Results Our meta-analysis of placental gene expression patterns revealed 419 differentially-regulated genes (136 downregulated and 283 upregulated) at false discovery rate(FDR)<0.05. Genes coding for TFIIH subunits and for components of RNA Pol-II complex were significantly-downregulated in preeclamptic placentas. Interrogation of the preeclampsia-specific gene signature through a filter constructed from the three relevant data sources identified EGFR and ATP3 as key regulators of preeclampsia development. ATP3 upregulation was found in an upstream event to TERT and Cri du Chat syndrome (OMIM#147791) and Canonical Cancer Institute, NIH, Frederick, MD.

Posters: Prenatal, Perinatal and Reproductive Genetics
2870T
TLR SNP T399I and early gestational age in a Wisconsin population of black newborn infants. D. Pillers1, M. Baker1, S. Schrod1, L. Zyduck1, J. DeValk1, B. Pattnaik1, S. Tokarz1, 1) Pediatrics, University of Wisconsin, Madison, WI, United States; 2) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI United States.

Background: Toll-like receptors (TLRs) are present in many cell types and serve as the first point of defense in the innate immune system by initiating the inflammatory cascade in response to infection. Single nucleotide polymorphisms (SNPs) are present in many TLR genes and have been associated with disorders of inflammation. TLR4 D299G and T399I SNPs are associated with increased susceptibility to infection from various pathogens. As inflammation can play a role in the development of preterm labor, TLR SNPs that alter the response to infection may have a different frequency in the preterm population. Objective: Screen a large, diverse population of Wisconsin newborns for TLR4 D299G and T399I SNPs and determine the genotype frequency in relation to gestational age (GA). Design/Methods: Anonymized DNA samples from 3095 infants were obtained in collaboration with the Wisconsin State Laboratory of Hygiene. Race was self-determined by patient as White (hispanic/non-hispanic), Black, Asian, and American Indian. TLR SNP assays (rs4986790, rs4986791) were purchased from Applied Biosystems Inc (ABI) and run on the ABI StepOnePlus™. Data were analyzed using StepOne™ software v2.2.2. Statistical analysis was carried out in collaboration with the Marshfield Clinic and the National Institute of Standards and Technology. Results: We did not find a significant correlation between the 299G and 399I alleles and early gestational age in the total population. However, when adjusting for racial background, we found that among Black infants the 399I allele frequency was 0% in term infants (37-41 weeks GA), whereas in preterm infants <37 weeks GA (P=0.039) and 7.7% in those <33 weeks GA (P=0.020). Conclusions: We show that in a population of Wisconsin infants the TLR4 SNP 299G did not associate with early gestational age. Our data are not consistent with current literature. We did not observe a correlation between the TLR4 SNP 299G allele and early gestational age. However, we did show an association between the 399I allele and early gestational age in a cohort of Black infants. Our data suggest that the racial background of an individual may affect the contribution of a specific TLR allele to preterm birth.

2872T
LncRNA Pathways Involved in Premature Preterm Rupture of Membrane (PPROM). Q. Shi1, X-C. Luo1, Y. Gu1, J. Pan1, N. Zhong1,2,3,4, 1) Lianyungang Maternal and Children's Hospital, Lianyungang, China; 2) March of Dimes Global Network of Maternal and Infant Health, NY; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY. Preterm birth (PTB) is a live birth delivered before 37 weeks of gestation (GW). About one-third of which result from the preterm premature rupture of membranes (PPROM). Presently, the pathogenic mechanism underlying PPROM is not yet clearly understood. In this study, we have investigated the dynamic expression of IncRNAs in placentas of PPROM, compared to controls, and their involvement in the pathogenic pathway of PPROM. A total number of 1954, 776, and 1050 IncRNAs were identified from placentas of PPROM (group A), which were compared to full term (FT) birth (group B), PTB (group C), and premature rupture of membrane (PRM) (group D), respectively. Instead of investigating the individual pathogenic function of each IncRNA involved in the pathogenic mechanism underlying PPROM, we have focused on investigating the metabolic pathways and their functions to explore what is the likely association and how they are possibly involved in the development of PPROM. Six groups, including up-regulation and down-regulation in the comparison of A vs. B, A vs. C, and A vs. D, of pathways were analyzed. Our results showed that 22 pathways were characterized as up-regulated 7 down-regulated in A vs. C, 18 up-regulated and 15 down-regulated in A vs. D, and 33 up-regulated and 7 down-regulated in A vs. B. Functional analysis showed pathways of infection and inflammatory response, ECM-receptor interactions, apoptosis, actin cytoskeleton, and smooth muscle contraction are the major pathogenic mechanisms involved in the development of PRM, which opens a new avenue for further investigating of the regulation of IncRNAs in PPROM as well as PTB.

2873W
Copy number variation of RYR1 locus, which is involved in myometrial contraction and relaxation, is associated with preterm births. M. Liu1, Y. Chen1,2, N. Zhong1,2, 1) Center for Reproduction and Genetics, Suzhou Municipal Hospital Affiliated to Nanjing Medical University, Suzhou, Jiangsu, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children's Health, China; 3) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY.

Preterm birth refers to a life birth before 37 gestational weeks. It is the leading cause of neonatal death. As well, it has been listed as one of the most common causes for under-5 children’s mortality and morbidity. Although many risk factors have been identified to associate with preterm birth, genetic deficit is yet unknown. In this study, we have applied a genome wide study to assess the association of copy number variation (CNV) with preterm birth. A cohort of 26 preterm birth samples with gestational age of <32 weeks, for which, common known factors had been excluded, were analyzed with Agilent 1M microarray comparative genomic hybridization (aCGH), compared to the results of 49 normal samples. 639 CNV loci were found from preterm birth samples and 12 CNVs, mainly located on chromosome 1, 17, 19, were associated with preterm birth. Among these 12 loci, three were shown statistically significant. Variation within these loci includes duplication, homozygous and heterozygous deletion. Among 26 preterm birth samples, 3 (12%) at three loci and 8 (31%) at two loci had copy number variations simultaneously. A gene RYR1 located within the significant loci were noticed to be involved in calcium regulation in smooth muscle and most interestingly, in myometrial contraction and relaxation that indicates this gene may contribute to the premature labor for preterm birth.
2874T
An Inverse Association Between Telomere Lengths and Gestational Age. H. Naden1, K. Ryckman2, J. Dagle3, J. Murray4. 1) Carver College of Medicine, University of Iowa, Iowa City, IA; 2) Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA; 3) Department of Pediatrics, University of Iowa, Iowa City, IA.

Preterm birth (PTB; defined as before 37 weeks of completed gestation) is a complex disorder and is associated with significant neonatal morbidity and mortality rates. Genetic variations between preterm and term infants may provide insights into the pathways involved with PTB. One such genetic variation of interest is telomere length (TL). In humans, telomeres are non-coding TTAGGG repeats located at the end of chromosomes. With each cell division, telomeres shorten. Eventually shortened telomeres lead to cell senescence and serve as one marker of aging. Shorter TL has been associated with cardiovascular disease, cancer, dementia, arthritis, osteoporosis, and hypertension. The main determinants of TL are age, inheritance, and gender. Studies have shown conflicting results regarding a relationship between gestational age (GA) and TL. We tested the association between TL and GA in an initial 260 cord blood samples taken from a cohort of infants with GA between 24-41 weeks. Relative TL was measured using a qPCR method. Primers specific to telomeres (the target of interest) and the 36B4 gene (a single copy gene on chromosome 12 used as a reference gene) were used to obtain a ratio of telomere to single copy gene length (T/S). This ratio was normalized to control for quantity of genomic DNA. Standard curves were produced as a means to check the efficiency of the assay. Relative quantification was used to analyze the data. The data was stratified into Group 1, GA less than 37 weeks (n=166), and Group 2, GA of 39 to 41 weeks (n=75). The average GA for the first group was 30.7 ± 3.7 and for the second group was 40.0 ± 1.7 weeks. The normalized T/S ratio of Group 1 is 3.69 while that of Group 2 is 2.18. This indicates that the TL of the preterm group (Group 1) is about 1.7 times longer than that of the term group (Group 2). A single variable regression of TL on GA yields a beta value of 0.083 (p-value = 0.033) indicating that a one unit increase in GA results in a TL loss of 0.093. The effect of TL on PTB outcome, as well as how TL changes with respect to gestational age, will now be examined with a larger sample size.

2875W
Case report of a pregnant woman with inherited thrombocytopenia associated with MYH9 mutation. O. Saµura1,2, M. Mizuno2, S. Kunishima2. 1) Dept OB/GYN, Onomichi General Hospital, Onomichi, Hiroshima, Japan; 2) Dept OB/GYN, Kure Medical Center, Kure, Hiroshima, Japan; 3) Dept of Advanced Diagnosis, Clinical Research Center, Nagoya Medical Center, Nagoya, Japan.

[Introduction] MYH9 disorders are autosomal dominant macrothrombocytopenia with leukocyte inclusion bodies, characterized by giant platelets, thrombocytopenia, and Döhle body-like cytoplasmic inclusion bodies. May-Hegglin anomaly (MHA), caused by mutations of MYH9, is classified as a MYH9 disorder. Here, we report a rare case of a pregnant woman with familial congenital thrombocytopenia and confirmed MHA genetic diagnosis. [Case] A 31-year-old woman was diagnosed with thrombocytopenia when she was 20-year-old, but she did not receive treatment. Her family history included low platelet levels in her paternal grandfather, father, and younger sister, and during her first pregnancy the platelet count was approximately 5.0 × 10^4/µL. However, she had a vaginal delivery without excessive bleeding. The patient was referred to our department for the management of her second pregnancy; however, due to her low platelet count (4.7 × 10^4/µL) and family history, congenital thrombocytopenia was suspected. A blood film revealed giant platelets, disparity in platelet size, and Döhle body-like cytoplasmic inclusion bodies, which were present in almost all segmented leukocytes. By immunofluorescence staining, abnormal myosin aggregation in granulocyte cytoplasm was suggested; therefore, we performed MYH9 testing with the consent of the patient and her father at gestational week 30. Genetic testing revealed an E1841K substitution in exon 38 of MYH9, and MHA was confirmed. Her platelet count ranged from 4-6 × 10^4/µL during her second pregnancy without the need for Boocytopenia. She delivered a normal baby after vaginal delivery of a 3148-g female with Apgar scores of 9 and 9 at 1 and 5 min, respectively, at gestational week 40. However, the second child also had thrombocytopenia and the same abnormal morphological findings as our patient. [Conclusion] In pregnancy, thrombocytopenia management with a family history should be carefully observed by peripheral blood smears because MYH9 disorders are not merely benign complications and should be closely monitored in pregnancy. The only known ligand for KIR2DL4 is HLA-G, which is highly expressed at the maternal-fetal interface and also conducts an important role during pregnancy. We previously examined the effect of HLA-G polymorphism on perinatal HIV-1 transmission and observed HLA-G*01:03 being associated with a decreased risk of transmission. In this study we analyzed the genetic polymorphism of KIR2DL4 and their epistatic interactions with HLA-G in perinatal HIV-1 transmission. Design: One hundred and ninety four HIV-1 positive drug naive pregnant women who delivered a child in a mother-child HIV-1 transmission cohort in Nairobi, Kenya were genotyped for 2DL4 using a sequence-based typing method. Associations of major 2DL4 alleles and their epistatic interactions with HLA-G alleles were analyzed by statistical analysis using SPSS 13.0 for 274 mother-child pairs. Results: A total of 20 2DL4 alleles were identified in this population, with 8 of them characterized. Only 6 alleles had a phenotype frequency of greater than 5%. 2DL4*0010301 was the most common allele with an allele frequency of 42.71% and phenotype frequency of 62.81%. 2DL4*022 was associated with an increased perinatal HIV-1 transmission (p=0.009, OR=2.977 95% CI:1.276-6.942). Epistatic interactions between 2DL4*008 and HLA-G*01:02:01 increased the odds of perinatal transmission. The odds ratio for mothers with HLA-G*01:02:01 alone was 1.322 (p=0.28, 95% CI:0.936-1.89) and for 2DL4*008 alone was 4.617 (95% CI:0.654-2.046). However, in mothers with both of these alleles the odds ratio of transmitting HIV-1 to their children increased to 2.011 (p=0.035, 95% CI:1.044-3.876). Conclusions: KIR2DL4 polymorphism and its interaction with HLA-G can influence the risk of perinatal HIV-1 transmission.
2879W
Provil loads of human T-cell leukemia virus type 1 in the peripheral blood samples from carrier pregnant women. N. Fuchi1, T. Tsukiyama2, T. Otsuki3, Y. Inokuchi4, K. Yanagihara5, S. Kanihira6, H. Moriiuchi7, K. Yoshiura8, H. Masuzaki1. 1) Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Medicine, Nagasaki, Japan; 2) Department of Laboratory Medicine, Nagasaki University Graduate School of Medicine, Nagasaki, Japan; 3) Department of Pediatrics, Nagasaki University Graduate School of Medicine, Nagasaki, Japan; 4) Department of Human Genetics, Nagasaki University Graduate School of Medicine, Nagasaki, Japan.

Objective: Human T-cell leukemia virus type 1 (HTLV-1) is a causative virus of adult T-cell leukemia (ATL), which is still intractable against any medical treatments. As the proviral loads of HTLV-1 in blood samples is associated with the risk of ATL later, those in pregnant women may be also a risk factor of maternal-child transmitted infection. However, the information regarding proviral loads of HTLV-1 in carrier pregnant women is still unknown. Therefore, the aim of this study is to clarify the changes of proviral loads of HTLV-1 genome in the peripheral blood samples from carrier pregnant women before and after delivery. Material and Methods: When the pregnant woman was decided as positive or false positive by the first screening test by chemiluminescent enzyme immunoassay (CLEIA), blood sampling before delivery was performed at 28-34 weeks of gestation and the quantitative real-time PCR of HTLV-1 genome (30ng genomic DNA) was done as confirmation test. When HTLV-1 genome was detected in blood sample, pregnant woman was diagnosed as HTLV-1 carrier. After delivery, blood samples of HTLV-1 carrier pregnant women were obtained within 24 hours, and the proviral loads of HTLV-1 were measured by quantitative real-time PCR before and after delivery. PCR primers for HTLV-1 genome were located at the pX region, while PCR primers for beta-globin as an internal control were on exon 2. HTLV-1 proviral load was calculated by the formula [(HTLV-1 pX copy number)/(beta-globin copy number/2)]×10,000 cells. Result: A total 28 pregnant women were diagnosed as HTLV-1 carrier. Mean (minimum-Maximum) proviral loads were 83.3 (0.87-502.3)×104 cells before delivery, while 34.27 (0-149.37)×104 cells after delivery. HTLV-1 proviral loads before delivery was higher than those after delivery (Wilcoxon signed rank test, P=0.01). Conclusion: The proviral load of HTLV-1 genome in blood samples was decreased significantly after delivery, suggesting that the changes of HTLV-1 proviral load may reflect the changes of CD4-positive T cells in pregnant women.

2880T

Objective: The aim of this study was to characterize placenta-specific microRNAs in fetal growth restriction (FGR) pregnancy. Method: Placenta-specific microRNAs were identified by next-generation sequencing analysis. Subsequently, quantitative real-time RT-PCR was used to identify FGR placenta-specific microRNAs whose level of expression was significantly decreased in FGR placenta (n=45) compared with uncomplicated placenta (n=50). FGR pregnancy-associated, placenta-specific microRNAs were identified in maternal plasma after delivery at significantly decreased concentrations, and their circulating levels in maternal plasma was compared between FGR pregnancies (n=10) and uncomplicated pregnancies (n=10). Results: Out of the 10 placenta-specific microRNAs that we identified, seven placenta-specific microRNAs (hsa-miR-518b, hsa-miR-1323, hsa-miR-516b, hsa-miR-515-5p, hsa-miR-520h, hsa-miR-519d and hsa-miR-526b) from the chromosome 19 microRNA cluster were identified as FGR placenta-specific microRNAs. Four FGR placenta-specific microRNAs (hsa-miR-518b, hsa-miR-1323, hsa-miR-520h and hsa-miR-519d) were confirmed as FGR pregnancy-associated, placenta-specific microRNAs, but their circulating levels in maternal plasma will be crucial to understanding how placenta-specific microRNAs are released into the maternal circulation.

2881W
Heritability of cardiovascular diseases in a preeclampsia family cohort. L.C.V. Thomesen1,2, P.E. Melton3, K. Tollaksen4, I. Lysbo5, P. Solberg6, L.T. Roten7, A.S. Gundersen6,7, M.L. Ofeldt8, K.M. Strand9, O.K. Nygård4,8, C.Sun10, A.C. Iversen11, R. Austgulen12, E.K. Moses13, L. Bjørges14,1. 1) Department of Obstetrics and Gynecology, Haukeland University Hospital; 2) Department of Clinical Science, University of Bergen; 3) Centre of Genetic Origins of Health and Disease, University of Western Australia; 4) Department of Obstetrics and Gynecology, Stavanger University Hospital; 5) Department of Obstetrics and Gynecology, Leverkusen Hospital; 6) Department of Laboratory Medicine, Children’s and Women’s Health, Faculty of Medicine, Norwegian University of Science and Technology; 7) Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology; 8) Department of Heart Disease, Haukeland University Hospital.

Purpose: Preeclampsia (PE) is a complex genetic disease of pregnancy with potentially severe outcomes for mother and fetus. Women who develop PE have been observed to have increased risk (2-8x) of developing cardiovascular diseases (CVD) later in life. We aimed to estimate the heritability of CVD phenotypes in well-characterized families with an acknowledged predisposition to PE. Methods: Using information contained in the Norwegian Preeclampsia Family Biobank, a cohort composed of families with an increased occurrence of PE, we classified 496 participants from 138 families according to predefined CVD phenotypes; a) risk factors for CVD (treated hypertension and/or hypercholesterolemia), b) established atherosclerotic vascular disease (subtypes: angina, acute myocardial infarction, stroke, aneurysm, thrombosis), or c) no diagnosis of CVD. Heritabilities were calculated using a variance components procedure implemented in the SOLAR software with the inclusion of covariate age, sex, and their interactions. Results: Average age of the participants was 46.9 (range 18-87) years. Heritability estimates (h²) were statistically significant for PE (h²=0.54±0.3, p=0.02), non-pregnancy related hypertension (h²=0.82±0.1, p=2.1×10-3) and hypercholesterolemia (h²=0.55±0.3, p=0.03). When only the presence or absence of any of the CVD phenotypes was assessed the heritability of CVD was highly significant (h²=0.31±0.12, p=1.9×10-3). The trait heritability was not significant for the other CVD phenotypes examined. Conclusion: Our study suggests that heredity may play a role in the development of PE, and specific CVD traits identified in a family cohort. Due to the relative young age of these participants most of the cohort would not yet have experienced manifest atherosclerotic disease in form of acute myocardial infarction or stroke. Further clarification of the established risk factors of CVD. Further identification of genetic variants specific to the affected families will allow us to focus on certain genetic areas potentially regulating biologic functions implicated in development of both CVD and PE.

2882T

Amniotic fluid is the only body fluid in direct contact with the fetal oropharynx, lungs, gastrointestinal tract, skin, and urinary system and amniotic fluid cell-free fetal RNA (AF cffRNA) can provide biological information on developing fetal organ system. The aim of this study is to clarify AF cffRNA profile of normal fetuses in three different gestational periods, 16–17 weeks (group 1), 25–26 weeks (group 2), and 36–38 weeks (group 3). The total 13 AF samples, 5 of group 1, 4 of group 2 and 4 of group 3, were collected. Total RNA was extracted from 10 ml of the amniotic fluid supernatant using the QIAamp Circulating Nucleic Acid Kit. The extracted RNA was hybridized to GeneChip® PrimeView™ Human Gene Expression Array to determine gene expression in amniotic fluid supernatant samples. The data were analyzed with Robust Multi-array Analysis using Affymetrix default analysis settings and global scaling as normalization method. The normalized, and log transformed intensity values were then analyzed using GeneSpring 12.5. In group 1, total 3269 transcripts were identified. We were conducting experiments, 532 transcripts have been known to be associated with fetal development through a search of pathway analysis software and Medline. We found 1,977 genes to be differentially expressed between group 1 and group 2. Among these genes, 27 genes increased and 1950 genes decreased in group 2. Also, we found 2,039 genes to be differentially expressed in group 3 compared to group 1. 48 out of 2,039 genes increased and 1475 genes decreased in group 3. Our result presented cffRNA profiles in amniotic fluids of pregnant women with normal fetal development in three different gestational periods. This study will help us in further studies for the discovery of biomarkers for abnormal fetal growth.
2883W
Recurrent enlarged nuchal translucency: prenatal presentation of a familial 15q26.2→qter deletion syndrome. R. Reiss1, D. Ahern1, M. Sandstrom2, L. Wilkins-Haug1. 1) Center for Fetal Medicine and Prenatal Genetics, Dept of Obstetrics and Gynecology, Brigham and Women’s Hospital, Boston, MA; 2) Dept of Pathology, Brigham and Women’s Hospital, Boston, MA.

Objective: To elucidate familial recurrence of enlarged nuchal translucency (NT) with chromosomal microarray and FISH techniques. Methods: A patient presented in each of 3 pregnancies with enlarged NT (4.7 mm, 4.0 mm, 7.1 mm). The first pregnancy was terminated when cardiac defect was suspected at 13.1 wks. Products of conception (POC) were karyotyped as 46,XX. In the next pregnancy, chorionic villus sampling at 12 wks showed karyotype 46,XX, and no deletion at 22q11.2. Ultrasound at 17 wks identified talipes. Fetal echocardiogram was normal at 19 wks. Pregnancy continued, complicated by IUGR. Multiple anomalies were identified after birth (micrognathia, malrotated kidney, talipes, brachydactyly). Postnatal growth lag prompted microarray testing at 5 months. In a 3rd pregnancy, septated NT at 11.7 wks again led to pregnancy termination. Chromosome analysis of POC showed 46,XY. Microarray was performed on fetus. Karyotype and FISH were performed on the mother. Results: 6.0 SNP microarray on the liveborn girl showed a 6.53Mb deletion at 15q26.2-q26.3, confirmed by 15q subtelomeric FISH, CGH microarray on fetal DNA from pregnancy 3 showed a 6.49Mb loss of 15q26.2→qter, consistent with the deletion seen in the affected child. FISH using a 15q subtelomere probe confirmed microarray findings. Subtelomeric 15q FISH analysis, performed retrospectively on a destained slide of cultured metaphases of POC from pregnancy 1, confirmed 15q26.2-26.3 deletion. Maternal blood chromosome analysis was normal, 46,XX. FISH using BAC probes from 15q26.2 and 15q26.3, with control probe to 15q11.2, showed no deletion or translocation of the 15q26.2 region. Studies on father and CVS from a recent 4th pregnancy are pending. Conclusions: Terminal microdeletion 15q26.2→qter, an emerging syndrome of pre and postnatal growth lag and anomalies, can present in 1st trimester with large NT. Our case highlights the importance of offering chromosomal microarray as well as conventional karyotyping when large NT is identified. Retrospective performance of microarray or FISH on DNA/cells from fixed samples can help document a familial syndrome. Based on review of the literature, most cases of terminal 15q deletion are de novo. This is the 2nd report of a familial case. We postulate that the father may carry a chromosomal rearrangement involving the 15q terminal region. Alternatively a small inversion in the region of the breakpoints, or germline mosaicism might be present in a parent.

2884T
Gene expression differences between preeclamptic and healthy placentas - an RNA sequencing study. T. Kaartokallio1, A. Cervera2, S. Hautaniemi2, J. Kere1,3,4,5, H. Laiuori1,6. 1) Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland; 2) Systems Biology Laboratory, Institute of Biomedicine and Genome-Scale Biology Research Program, University of Helsinki, Helsinki, Finland; 3) Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden; 4) Molecular Neurology Research Program, Research Programs Unit, University of Helsinki, Finland; 5) Folkhälso Institute of Genetics, University of Helsinki, Helsinki, Finland; 6) Department of Obstetrics and Gynaecology, Helsinki University Central Hospital, Helsinki, Finland.

Objectives: Preeclampsia (PE), a common pregnancy disorder originating from impaired placental development, is responsible for morbidity and mortality to both mother and child. In the present study, we utilized RNA sequencing to identify genes that are differentially expressed between preeclamptic and healthy placentas. Methods: Placental RNA of nine preeclamptic women and nine healthy controls was sequenced in pools of three with Illumina chemistry. The PE pools 1, 2 and 3 consisted of placentals samples from 38-39, 34-36 and 33 weeks of gestation, respectively, and the control pools from 38-39 weeks of gestation. All women delivered by C-section without labor. The data was processed and analyzed with appropriate tools including Tophat, Cuffdiff and CummeRbund. The two groups were tested for differential gene expression by comparing all the case pools as well as subsets of the case pools against the control pools. The results will be validated by qPCR in a sample set of 20 cases and 19 controls including the samples utilized in the RNA sequencing. In this sample set, the cases and controls do not differ significantly for maternal age or body mass index, gestational age, mode of delivery or sex of the child. Results: Initially, a total of 58 genes were found to be differentially expressed between PE and control groups. After excluding genes with inconsistent expression pattern within the groups, hemoglobin genes, and genes located in Y chromosome, as the fetal sex distribution between the groups was not equal, we were left with 40 genes. Of these, 15 will be subsequently validated by qPCR. The genes selected for the validation are involved e.g. in immunological processes, placental development and function, invasion, angiogenesis and vasodilation, all of which are processes relevant for the development of PE. Some of these genes have also been previously associated with autoimmune or renal diseases. In addition to our main goal of identifying expression differences between preeclamptic and healthy placenta, we wish to emphasize the importance of taking account of gestational age when studying placental gene expression in pregnancy complications. Therefore, in the qPCR validation, we also aim to demonstrate the effect of gestational age on expression of certain genes. Conclusions: By comparing transcriptomes of preeclamptic and healthy placentas, we were able to identify genes that were differentially expressed between the two groups.
2885W Toward a Mouse Model of Thrombocytopenia with Absent Radius (TAR) Syndrome. V.L. Homer1, A. Doddi1, A. Long1, C.L. Martin2, T. Casper2. 1) Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Autism and Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA.

TAR syndrome is characterized by low platelet counts (thrombocytopenia) and absence of the radius bone in each forearm. Thrombocytopenia can lead to easy bruising and/or hemorrhage in infancy, which may become less severe over time. The absence of radius bones results in shortened forearms. TAR syndrome is inherited in an autosomal recessive manner and is due to mutations in the gene huntingtin (Htt). In the majority of HD patients, an expanded CAG repeat causes the development of HD. The Htt mutant HD gene can be expressed from a single expanded allele. To increase the chance of identifying early responses to therapy, we have generated mouse alleles in which the endogenous RBM8A gene is under the control of a tetacycline-dependent transactivator. This system will enable us to control the amount of Y14 protein available during embryonic development, by varying the amount of tetracycline fed to pregnant mothers. We are generating two strains of mice: the first strain will have the transactivator (tetO) inserted in the endogenous RBM8A gene. In this strain, the transactivator will drive the expression of endogenous RBM8A. The second strain will have the tetracycline-responsive promoter (tetO) driving RBM8A expression. The presence of tetracycline allows the transactivator to bind and drive RBM8A transcription. There is a linear relationship between the amount of tetracycline and the amount of Y14 protein synthesized. This system allows us to examine the temporal and spatial requirements for RBM8A during development. This will enable us to control the amount of Y14 protein during embryonic development and by driving the transactivator with promoters specific for various tissues. These experiments are the first step in defining the therapeutic level of Y14 for TAR syndrome treatment.

2886T A single mutant HTT allele is sufficient to elicit early alterations to the brain in a knock-in mouse model of Huntington’s disease. M. Kovalenko1, A.J. Milnerwood1, S. Tappan2, J. St Claire1, J.R. Guide1, R.C. Switzer3, L.A. Raymond1, J.-M. Lee1, V.C. Wheeler1. 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 2) Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada; 3) Microbrite(TM) laboratories, MBF Labs, Williston VT, USA.

Huntington’s disease (HD) is a dominantly inherited neurodegenerative disease caused by the expansion of a CAG repeat in the HTT gene encoding huntingtin (Htt). To increase the chance of identifying early responses to treatments relevant to the disease process in patients, we are using accurate genetic knock-in mouse models of HD to delineate the critical dosage threshold of Y14 that causes TAR syndrome phenotypes as well as the dosage that causes lethality. Ultimately, this system will allow us to control the amount of Y14 protein available during embryonic development, by varying the amount of tetracycline fed to pregnant mothers. We are generating two strains of mice: the first strain will have the transactivator (tetO) inserted in the endogenous RBM8A gene. In this strain, the transactivator will drive the expression of endogenous RBM8A. The second strain will have the tetracycline-responsive promoter (tetO) driving RBM8A expression. The presence of tetracycline allows the transactivator to bind and drive RBM8A transcription. There is a linear relationship between the amount of tetracycline and the amount of Y14 protein synthesized. This system allows us to examine the temporal and spatial requirements for RBM8A during development. This will enable us to control the amount of Y14 protein during embryonic development and by driving the transactivator with promoters specific for various tissues. These experiments are the first step in defining the therapeutic level of Y14 for TAR syndrome treatment.


Orofacial defects such as cleft lip and palate (CLP) are the most common structural congenital anomalies, affecting 1 out of 700 births. Elucidating the genetic basis of orofacial clefts is essential to identify risk loci and develop new therapeutic and preventive measures. In an ongoing gene discovery effort, the Developmental Genome Anatomy project (DGAP) has developed whole genome sequencing strategies to characterize genes contributing to such human congenital anomalies. Through this methodology, we discovered a new candidate gene for CLP, ATG4C which encodes an autophagy related cytoskeletal protein. The isolated deletion of ATG4C was identified in a 12 year old Caucasian boy presenting a CLP. Previous studies in our lab have demonstrated conservation of the fundamental molecular pathways and homology of primary palate formation between mammals and zebrafish. Therefore we utilized the model organism zebrafish to study and validate the role of ATG4C in palate development. By whole mount in situ hybridization, we determined that the spatiotemporal expression of atgc is compatible with a role in craniofacial development. Additionally, knockdown studies performed with morpholin injections resulted in both a shortened and cleft palate, suggesting a strong role for atgc as the causal gene in the human case of CLP. Our ongoing effort is focusing on a mechanistic analysis of ATG4C in craniofacial morphogenesis and autophagy. Our future studies will track migrating cranial neural crest cells and help define if the failure of upper jaw morphogenesis in atgc mutants is due to a defect in cell migration, proliferation or convergence extension mechanisms. We also plan to explore a potential functional annotation of ATG4C to include detailed biological analysis of gene function, and highlights the utility of the zebrafish model in high throughput and reverse genetics approaches that are important in fundamental genomics projects.

2888W Trim37-deficient mice recapitulate several features of the multi-organ disorder Mulibrey nanism. K.M. Kettunen1, 2, 3, R. Karikoski4, R.H. Hannila1, T-T. Toivonen1, 2, 3, J.-M. Lee1, 2, 3, 5, L. Jalanko2, 4, A-E. Lehesjoki1, 2, 3, 5. 1) Folkhälsan Institute of Genetics, Helsinki, Finland; 2) Research Programs Unit, Molecular Neurology, Biomedicum Helsinki, University of Helsinki, Finland; 3) Neuroscience Center, University of Helsinki, Finland; 4) Central Hospital of Tavastia, Hämeenlinna, Finland; 5) Hospital for Children and Adolescents, Helsinki University Central Hospital.

Objective: Mulibrey nanism (MUL) is a rare autosomal recessive multi-organ disorder caused by mutations in the TRIM37 gene. The main symptoms include growth restriction, cardiopathy, infertility, increased risk for tumours and various skin disorders. MUL also shares features of the human disease phenotype, including growth restriction, infertility, risk of various tumors, cardiomyopathy, disturbances in glucose and lipid metabolism and hearing loss. MUL is considered an excellent model to study disease pathogenesis related to TRIM37 deficiency.

Towards a Mouse Model of Thrombocytopenia with Absent Radius (TAR) Syndrome. K.M. Kettunen1, 2, 3, R. Karikoski4, R.H. Hannila1, T-T. Toivonen1, 2, 3, J.-M. Lee1, 2, 3, 5, L. Jalanko2, 4, A-E. Lehesjoki1, 2, 3, 5. 1) Folkhälsan Institute of Genetics, Helsinki, Finland; 2) Research Programs Unit, Molecular Neurology, Biomedicum Helsinki, University of Helsinki, Finland; 3) Neuroscience Center, University of Helsinki, Finland; 4) Central Hospital of Tavastia, Hämeenlinna, Finland; 5) Hospital for Children and Adolescents, Helsinki University Central Hospital.

Objective: Mulibrey nanism (MUL) is a rare autosomal recessive multi-organ disorder caused by mutations in the TRIM37 gene. The main symptoms include growth restriction, cardiopathy, infertility, increased risk for tumours and various skin disorders. MUL also shares features of the human disease phenotype, including growth restriction, infertility, risk of various tumors, cardiomyopathy, disturbances in glucose and lipid metabolism and hearing loss. MUL is considered an excellent model to study disease pathogenesis related to TRIM37 deficiency.

Trim37-deficient mice recapitulate several features of the multi-organ disorder Mulibrey nanism. K.M. Kettunen1, 2, 3, R. Karikoski4, R.H. Hannila1, T-T. Toivonen1, 2, 3, J.-M. Lee1, 2, 3, 5, L. Jalanko2, 4, A-E. Lehesjoki1, 2, 3, 5. 1) Folkhälsan Institute of Genetics, Helsinki, Finland; 2) Research Programs Unit, Molecular Neurology, Biomedicum Helsinki, University of Helsinki, Finland; 3) Neuroscience Center, University of Helsinki, Finland; 4) Central Hospital of Tavastia, Hämeenlinna, Finland; 5) Hospital for Children and Adolescents, Helsinki University Central Hospital.

Objective: Mulibrey nanism (MUL) is a rare autosomal recessive multi-organ disorder caused by mutations in the TRIM37 gene. The main symptoms include growth restriction, cardiopathy, infertility, increased risk for tumours and various skin disorders. MUL also shares features of the human disease phenotype, including growth restriction, infertility, risk of various tumors, cardiomyopathy, disturbances in glucose and lipid metabolism and hearing loss. MUL is considered an excellent model to study disease pathogenesis related to TRIM37 deficiency.
2889T
An in vivo mouse model to study the phosphorylation of FMRP. M. Santoro, S.T. Warren. Department of Human Genetics Emory University Atlanta, GA.
Frágil X syndrome, the most common form of inherited intellectual disability, results from a lack of Fragile X Mental Retardation Protein (FMRP). FMRP is a selective mRNA-binding protein and plays a key role in synaptic plasticity. In neurons, FMRP represses translation of its target mRNAs but allows translation to proceed upon receipt of specific synaptic signals. In the hypothalamus, activation of mGluR receptors causes FMRP to de-repress translation of its target mRNAs, resulting in a burst of local protein synthesis, internalization of AMPA receptors, and long term depression. In vitro experiments indicate that FMRP’s activity is regulated by the phosphorylation status of a specific serine (S499 in mice; S500 in humans). To explore this, we have developed a knock-in mouse model with a Ser499Ala mutation. This results in an amino acid replacement that structurally resembles unphosphorylated FMRP. Indeed, using an antibody raised against phospho-FMRP, no signal is observed on Western blots from Ser499Ala mouse tissue, although normal levels of FMRP are seen with the conventional antibody. Phenotypic studies are underway but preliminary data suggest that the Ser499Ala mouse may not be a complete phenocopy of the FMRP null mouse.

2890F
A systematic genome-wide knockout generation and analysis of zebrafish protein-coding gene function. C.M. Dooley, E.M. Busch-Nentwich, R.N.W. Kettleborough, C. Scahill, I. Sealy, R.J. White, J.C. Collins, N. Wall, C. Herd, R. Gibbons, S. Carruthers, A. Halli, R.C. Clark, Z. Pusztai, M. Niemi, F. van Eeden, J.C. Barrett, D.L. Stemple. 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) MRC-Crystal Informatics, University of Sheffield, Sheffield, United Kingdom. Although the human genome sequence was completed a decade ago, less than half of the identified vertebrate genes have been assigned a function. Detailed investigations involving model organisms have played a fundamental role in connecting genotype to phenotype and will continue to do so. The ability to thoroughly investigate the role of a gene in various biological processes greatly depends on loss of function analysis which has traditionally been carried out, in vertebrates, on a gene-by-gene basis. The zebrafish is a genetically tractable organism with an external, transparent and rapidly developing embryo offering an in toto model of the complex cellular processes required for the formation of all major vertebrate organs. In recent years it has moved beyond being a model of early development as 70% of all human genes are represented by at least one zebrafish ortholog. This has helped to make zebrafish a first stop choice for human geneticists aiming to assign function to human orthologs or clinicians hoping to develop a model for a particular disease. Enabled by a completed and annotated zebrafish reference genome sequence, high-throughput sequencing and efficient chemical mutagenesis, we describe the active Zebrafish Mutation Project (ZMP) at the Wellcome Trust Sanger Institute which aims to identify and phenotype disruptive mutations in every zebrafish protein-coding gene. Thus far we have identified potentially disruptive mutations in more than 48% of all of the 26,000 known zebrafish protein coding genes. We have developed a multi-allelic phenotyping scheme to efficiently assess the effects of each allele during embryogenesis and have analyzed the phenotypic consequences of more than 1600 alleles. Our phenotyping scheme is also adaptable to phenotypic analysis beyond embryogenesis and we are currently expanding our analysis to encompass a wider array of phenotypes as we set forth to fully functionally annotate the zebrafish genome. To accompany our chemical mutagenesis we have also developed the Cas9/CRISPR system to allow for single as well as multiplexed targeted bespoke knockouts. A further goal is to use our rich collection of knockouts and knockout generation capabilities to model human disease and to work towards the complete functional annotation of the entire zebrafish genome. To this end, all knockout alleles and data are immediately made available to the community via our website http://www.sanger.ac.uk/cgi-bin/Projects/D_relio/zmp.

2891W
The laboratory mouse is the primary model for human disease based on its extensive genetics, fully-sequenced genome, and large-scale mutagenesis programs that have created ENU (point mutations) and knock-out (null) mutations covering most of its genome. In addition, programs to develop genetically-defined variant populations mimicking human population variation provide new means to study quantitative traits and complex inherited syndromes. This genomic knowledge, coupled with increased phenotyping power and micro-technologies, adds exquisite detail to knowledge about morphological and physiological variation in mouse.
The Mouse Genome Informatics (MGI, www.informatics.jax.org) database and resources provide a roadmap to discovery for biologists and clinicians seeking to use the extensive genetic, phenotypic, and disease model data available to correlate mouse phenotypes and human disease. MGI holds data on more than 750,000 mutant allelic variants in 21,400 genes. More than 250,000 phenotypic annotations to genotypes are curated using the Mammalian Phenotype Ontology. Mouse models are annotated to over 1,240 unique human diseases in OMIM. MGI is developing new interfaces to human disease information to foster use by clinical and human genetics researchers. These interfaces will provide easy access to mouse models via human genes, human genome locations, human disease symptoms, and human anatomy terms. Our aim is to reduce the barriers to use of mouse data and increase the synergy derived from comparative genomics and comparative phenotyping. This will facilitate mining data for common pathways, disease variants, and potential therapeutic interventions. We describe these new human-centric entries points to MGI data and present examples of insights derived from examining human diseases in light of mouse models. We demonstrate the ease of obtaining mouse disease models from repositories, the path to reference collections, and the search for collaborators. Supported by NIH grant HG003303.

2892T
Reducing variability in variation data using a common genotype-phenotype model. N. L. Washington, J. Brusht, A. Bandrowski, C. Boro- metc, K. Elbeek, J. Espino, J. Grethe, A. Gupta, H. Hochheiser, S. Hoffmann, J. Lewis, L. Liu, M. Martone, C. Mungall, P. Robinson, D. Smedley, C. Torniai, V. Vasilevsky, M. A. Haendel. 1) Genomics Division, Lawrence Berkeley National Lab, Berkeley, CA; 2) Library and Dept. of Medical Informatics and Epidemiology, Oregon Health Sciences University, Portland, CA; 3) Center for Research in Biological Systems, University of California San Diego, La Jolla, CA; 4) Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA; 5) Institut für Medizin- ische Genetik und Humangenetik, Charité - Universitätsmedizin Berlin, Berlin, Germany; 6) Wellcome Trust Sanger Institute, Cambridge, UK; 7) University of Utah, UT.
Investigating the basis of disease requires correlating and integrating geno- type, phenotype, and environmental factors. Ideally, we could leverage the depth of knowledge available from model systems. However, choosing an appropriate and sufficient model organism (e.g. genotype) for the study of a disease is a challenge, as it often requires an understanding of the organism’s development, normal mouse phenotypes, and a nuanced understanding of how traits are attributed to laboratory animal husbandry, breeding, and maintenance. Though the volume of curated genotype-to-phenotype data is steadily growing, our ability to query across different organisms and data resources is limited. This is due, in part, to a lack of a common genotype and phenotype reporting format, and results from the variety of ways in which this data is reported, including free text, organism-specific vocabularies, nomenclatures, syntaxes. The data is further complicated because phenotypes may be limited to specific aspects of a genotype, such as an allele, a QTL, or the complete genome. Despite efforts to standardize human variation through proposed standards such as HGVS many alternatives still exist, leading to difficulties when integrating human phenotype data from different database resources. A first step toward the integration of organism genotypes (including human and model systems) is to translate them to a common genotype model. As part of the Monarch Initiative, we have created a pan-species genotype representation that allows users to access human variation and model organism genotype-to-phenotype data in a uniform way. This permits cross-species querying by human diseases and nomenclature, and letting you know about a specific resources’ nomenclature, and is the first step in being able to aggregate and attribute phenotype data to different components of a genotype. This standardization enables translational genotype-phenotype hypotheses and analysis of treatment options, and will make model system data available for rare and undiagnosed disease diagnosis and clinical decision support tools.
2893F
MYBPC1 mutations impair skeletal muscle function in zebrafish models of arthrogryposis. D.M. Alvarado1, K. Ha1, J.G. Buchanan2, K. McCaill1, A. Vydyanchit1, P.K. Luther1, M.I. Goldsmith4, M.B. Dobbs1,3, C.A. Gurnett1,2,4 1) Orthopaedic Surgery, Washington Univ, St Louis, MO; 2) Neurology, Washington Univ, St. Louis, MO; 3) Heart and Lung Institute, Imperial College, London, UK; 4) Pediatrics, Washington Univ, St. Louis, MO; 5) St Louis Shriners Hospital for Children, St. Louis, MO.

Myosin binding protein C1 (MYBPC1) is an abundant skeletal muscle protein that is expressed predominantly in slow twitch muscle fibers. Human MYBPC1 mutations are associated with distal arthrogryposis type 1 and lethal congenital contracture syndrome (LCCS4). Because the function of MYBPC1 is incompletely understood, the mechanism by which human mutations result in contractures is unknown. Here, we demonstrate that mybpc1 is required for embryonic motor activity and survival in a zebrafish model of arthrogryposis using antisense morpholino knockdown. Mybpc1 morphant embryos have severe body curvature, cardiac edema, impaired motor excitation and are delayed in hatching. Myofibril organization is selectively impaired in slow skeletal muscle and sarcomere numbers are greatly reduced in mybpc1 knockdown embryos. To evaluate the effects of human distal arthrogryposis mutations, mybpc1 mRNAs containing the corresponding human W236R and Y856H MYBPC1 mutations were injected into embryos. Dominant-negative effects of these mutations were suggested by the resultant mild bent body curvature, decreased motor activity, myofibril disorganization, as well as impaired overall survival compared to overexpression of wild-type RNA. These results demonstrate a critical role for mybpc1 in slow skeletal muscle development and establish zebrafish as a tractable model of human distal arthrogryposis.

2894W
Scaling up: Integrating high throughput mouse phenotyping data with additional genomic resources for gene discovery. C.L. Smith, H. Dene, R. Balderelli, S. Gianatto, K. Forthover, J. Kadin, J. Richardson, J.T. Eppig, Mouse Genome Informatics Staff. Mouse Genome Informatics, Jackson Laboratory, Bar Harbor, ME.

The mouse is the most commonly studied model organism for understanding gene function in human disease and development. Comparative genetics and genomics can assist in prioritization of candidate genes for human disease, and model organisms can be used for confirmation or validation of gene function. However, comprehensive functional and phenotypic annotation of mouse genes is incomplete in many instances and mouse phenotype data are available for only 7770 of of the estimated 33,735 protein coding and RNA only coding genes. The International Mouse Phenotyping Consortium (IMPC) is gearing up to collect phenotype data on mice made from the large collection of knockout alleles in embryonic stem cells available from the International Knockout Mouse Consortium. The goal is to provide standardized phenotypic data derived from the same battery of tests for every gene in the mouse genome. Mouse Genome Informatics (MGI: www.informatics.jax.org) has already established pipelines to the WTSI and Europherome pilot phenotyping projects and has developed the infrastructure to integrate and display these initial sets of data with all of the other genetic and functional information available about these genes in mouse.

Phenotypic representation of genetically engineered mice integrated with other biological data at MGI will aid in model building, understanding biochemical pathways, and determining the underlying mechanisms of human genetic disease. We will show web based query results utilizing query forms designed to search phenotypic, biochemical function and process, subcellular location, expression, pathways and sequence data that may assist in finding novel candidate genes for human disease. Supported by NIH grant HG000330.

2895T
Translational Modeling of Calpain-5 Vitreoretinopathy Mechanisms in Mice. K.J. Wert1, 2, J.M. Skellet1, S.H. Tsang1, 2, V.B. Mahajan3,4 1) Department of Ophthalmology, Columbia University Medical Center, New York, NY; 2) Institute of Human Nutrition, College of Physicians and Surgeons, Columbia University, New York, NY; 3) Omics Laboratory, University of Iowa, Iowa City, IA; 4) Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA.

A wide range of human diseases including cancer, multiple sclerosis, Alzheimer’s disease, cataract, diabetes, and muscular dystrophy have implicated a role for calpains in their pathogenesis. We previously identified CAPN5 (human, hCAPN5) as the causative gene for retinal auto-immune uveitis in human patients with Autosomal Dominant Neovascular Inflammatory Vitreoretinopathy (ADNV). To understand the molecular and genetic mechanisms of ADNV, a human CAPN5 disease allele was expressed in mouse retinas using a lentiviral vector. Mouse Capn5 mRNA sequence, RNA expression, and protein expression was compared to hCAPN5 in the human retina. Lentiviral vectors were created to express either the wild-type human (h) CAPN5 or the ADNV mutant hCAPN5-R243L allele under a rhodopsin promoter with tandem green fluorescent protein (GFP) expression. Vectors were injected into the subretinal space of perinatal mice, and expression was determined by autofluorescent imaging. Mouse phenotypes were analyzed using electrophoretography, histology, and inflammatory gene expression profiling. Mouse calpain-5 shows high homology to its human ortholog with over 98 percent sequence identity that includes the ADNV mutant residue. In correlation with human expression, Capn5 RNA was detected in mouse retina. Calpain-5 protein was expressed in the inner and outer segments of both rod and cone photoreceptors and in the inner plexiform layer. Live imaging of the lentiviral GFP reporter showed good uptake and distribution throughout the retina. Electroretinography, retinal histology, and expression of inflammatory genes indicate the CAPN5-R243L allele elicits an ADNV-like disease in mice. Taken together, our studies suggest that ADNV is due to a CAPN5 gain-of-function, and retinal expression may be sufficient to generate an autoimmune response. Moreover, genetic models of ADNV can be developed in the mouse and serve as pre-clinical models for therapeutic testing. Further elucidation of the pathogenic mechanism of CAPN5 in ADNV, by examining the effects of hCAPN5 compared with hCAPN5 including the single amino acid change R243L, will provide important new insight into causes of irreversible human blindness.

2896F

The p.Arg345Trp mutation in the EFEMP1 gene causes the inherited macular degeneration Doyne Honeycomb Retinal Dystrophy/ Malattia Leventinese (DHRD/ML). The hallmark of DHRD/ML is the formation of drusen at an early stage. Gene targeted Efemp1R345W/R345W knockin mice develop extensive basal deposits which are considered precursors to drusen. In previous studies using Efemp1R345W/R345W;complement C3-/- double mutant mice we demonstrated a critical role for complement system in the formation of the basal deposits. The purpose of this study was to use a primary RPE cell culture model to investigate the mechanisms involved in the pathogenesis of basal deposit formation in the Efemp1R345W/R345W;complement C3-/- double mutant mice and the role of the complement system in this process. Primary RPE cells from wild type, Efemp1R345W/R345W knockin, and Efemp1R345W/R345W;C3-/- double mutant mice were grown on a permeable support under polarizing conditions. Cell phenotype was characterized by electron and fluorescence microscopy. Expression levels of complement genes and Efemp1 were quantified by qRT-PCR. Protein expression in the RPE cultures was determined by western blot analysis and immunofluorescence using confocal microscopy. As demonstrated by results of qRT-PCR experiments RPE cells from each strain express the following complement components and factors: C1q, C2, C3, C4, C6, C8b, CI4 and C8f. Moreover, the expression of the C3 was significantly higher in Efemp1R345W/R345W knockin cells (ANOVA, p=0.0218) than in wild type. Electron micrographs showed that RPE cells from Efemp1R345W/R345W knockin but not wild type or double mutant Efemp1R345W/R345W;C3-/- mice secrete deposit-like material in vitro. Immunostaining of the deposits revealed that EFEMP1, C3 and other complement components and factors localize in the deposits. The primary RPE culture recapitulates in vivo observations. The primary RPE cell culture model is capable of reproducing deposit formation in vitro, which provides an excellent approach to investigate the very early effects of the Efemp1 mutation on RPE cell culture. The results suggest that the complement system is important and is essential in basal deposit formation in Efemp1R345W/R345W;C3-/- mutant mice. This model may be a useful system to test drugs, such as complement inhibitors, to prevent basal deposit formation in inherited macular degeneration as well as the more common age-related macular degeneration AMD.
2897W
Inactivation of the mir-183/96/182 cluster genes results in syndromic retinal degeneration. S. Xu1,2, S. Lumayag1,2,3, C.E. Haidi1,2,2, N.J. Corbett2, K.J. Watelin4, C. Cowan1,2,3, S. Turturro1,2,3, P.E. Larsen5, P.D. Witmer6, D. Valle7, D.J. Zack8, D.A. Nicholson9, 1) Pharmacology; 2) Ophthalmology; 3) Neurological Sci, Rush Univ Med Ctr, Chicago, IL; 4) Wilmer Eye Institute, The Johns Hopkins University School of Medicine, Baltimore, MD; 5) Biosciences Division, Argonne National Laboratory, Lemont, IL; 6) McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD.

The mir-183/96/182 cluster is highly expressed in the retina and other sensory organs. To uncover its in vivo functions in the retina, we generated a knockout mouse model, designated as mir-183-5’tg/tg, using a gene-trap embryonic stem cell clone. Our results showed that inactivation of the cluster resulted in early-onset and progressive synaptic defects of the photoreceptors, leading to abnormalities of scotopic and photopic ERGs with decreased b-wave amplitude as the primary defect and progressive retinal degeneration. In addition, inactivation of the mir-183/96/182 cluster resulted in global gene-expression changes in the retina, with enrichment of genes important for synaptogenesis, synaptic transmission, photoreceptor morphogenesis and phototransduction, suggesting that the mir-183/96/182 cluster plays important roles in postnatal functional differentiation and synaptic connectivity of photoreceptors. Our data suggest that the mir-183/96/182 cluster is a new candidate gene for inherited retinal degeneration and susceptibility gene for age-related retinal dysfunction/regeneration. Although the phenotypes and molecular changes in mir-183/96/182 mice are not a perfect match for any known human disease, mutation-negative Usher syndrome-like patients would be good candidates for mir-183/96/182 loss-of-function mutations. Studies on genetic variation or polymorphisms around the mir-183/96/182 gene are warranted in patients with age-related, progressive retinal and/or multiple sensory defects, and other neurological conditions.

2898T
The Dhkd1 Tyr486* knock-in mouse recapitulates some phenotypes of Charcot-Marie-Tooth disease type 2Q. M. Gu1, W. Guo1, C. Luan1, Z. Yu1, Y. Chen1, S. Dang1, Y. Kuang1, Z. Wang1,2,2, 1) Department of Medical Genetics, Shanghai Jiao Tong University School of Medicine (SJTUSM), Shanghai 200025, China; 2) Research Center for Experimental Medicine, Rui-Jin Hospital at SJTUSM, Shanghai 200025, China; 3) Shanghai Research Centre for Model Organisms, Shanghai 201203, China.

Charcot-Marie-Tooth disease (CMT) or peroneal muscular atrophy is one of the most common inherited neurological disorders with a prevalence estimated at 1/2500. On the basis of clinical manifestation and electrophysiological properties, CMT has been divided into four main types: CMT1, CMT2, CMT4 and CMTX. CMT2 is an axonal peripheral neuropathy characterized by distal muscle weakness and atrophy, normal or near-normal nerve conduction velocities. Our previous study demonstrated that the nonsense mutation [c.1455T→G (p.Tyr485*)) in exon 8 of DHTKD1 (dehydrogenase E1 class family like 1A) is a strong genetic factor for DHTKD1-related CMT2 (AJHG, 2012, 91: 1088-1094). To identify the molecular mechanisms of mutant DHTKD1 causing CMT2Q, we generated a knock-in mouse model for the DHTKD1 Tyr486* knock-in mouse model (Dhkd1-Tyr486*). At first, we evaluated the gene expression, energy production and apoptosis of the mutant Dhkd1 gene in the positive ES cell clones. We found that the gene expression and ATP production were significantly decreased in the heterozygous mice. Moreover, mutant Dhkd1 mice were similar to wild-type mice under hydrogen peroxide (H2O2)-induced oxidative stress. Then, the phenotypes were observed among wild-type mice, heterozygous mice and homozygous mice from two heterozygous parental mice. Taken together, we indicated that the Dhkd1 Tyr486* knock-in mouse model recapitulated some phenotypes of CMT2Q. Further in-depth studies are underway.

2899F
Comprehensive characterization of a zebrafish model for pseudoxanthoma elasticum reveals a role for the abcc6 transporter in cardiovascular development. M.J. Hosen1,2, O. Vanakker1, A. Wiltein1, P. Coucke1, A. De Paepe1, 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Dept. of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet, Bangladesh.

Pseudoxanthoma elasticum (PXE) is characterized by elastic fiber calcification and fragmentation, resulting from ABC6 mutations. Recently, up-regulation of the pro-osteogenic BMP2-RUNX2 pathway was demonstrated in patients. Also, recent reports on vascular malformations in patients suggested a possible developmental role for ABC6C. We used 4 zebrafish (ZF) models, targeting 3 abcc6 isoforms and combined knock-down, to assess the developmental phenotype, mineralization and mineralization-associated pathways. In-situ hybridization demonstrated all isoforms to be mainly expressed in the pronephric ducts. Abcc6 knock-down was obtained by injecting translational blocking or splice junction morpholinos (MOs) of abcc6 in 1-4 cell stage embryos. To avoid non-target-related phenotypes due to apoptosis, co-injection of an anti-p53 MO along with the experimental MO was performed. Morphants exhibited a delay in gastrulation. At 3 days post fertilization (dpf), curving and shortening of the tail, variable in severity, pericardial edema, decreased mobility, total body length and body pigmentation as well as underdevelopment of head and eyes were observed. Cardiovascular evaluation, using MO injection on flg/flg embryos demonstrated the development of abnormal rudimentary tubular heart with decreased heart beat and blood flow compared to controls, contributing to early demise of the morphants. Further, underdevelopment and delayed sprouting of multiple vessels was noted. Calcein staining of the morphants (5dpf) showed advanced skeletal mineralization compared to controls. QPCR analysis revealed up-regulation of bmp2a, runx2a and mscx. We illustrate a distinct phenotype, affecting longitudinal growth, eye and cardiovascular development. The cardiovascular underdevelopment in the morphants corroborates the hypothesis that ABC6C may play a role in PCP-mediated mineralization and confirmation of BMP2-RUNX2 involvement suggest that ZF is a useful model organism for further PXE cell signaling research.
2901T
Mutations in Sonic hedgehog signaling pathway predispose to fatty liver. A.F. Martinez1, R.J. Lipinski2, M. Guillen-Sacoto3, S.K. Hong1, J.L. Everson1, J. Soll2, R. Trichet1, A. inoue1, and D. Muenke1. 1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Dept. of Comparative Biosciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI; 3) The Bowles Center for Alcohol Studies, University of North Carolina, Chapel Hill, NC; 4) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Holoprosencephaly (HPH) is a severe genetic disorder that affects forebrain and craniofacial development. Several developmental factors implicated in HPH belong to or interact with Sonic hedgehog (Shh) signaling as the central player. In this demonstration, we have observed a significantly higher prevalence of fatty liver in HPH patients and unaffected carrier relatives (pediatric: p=0.0023, adult: p=0.0135) compared to the general population. While typically quiescent in the adult healthy liver, the Shh pathway becomes active in alcoholic and non-alcoholic fatty liver disease. The gene-environment studies described here explored the effect of Shh pathway mutations on liver metabolism using a mouse model of fatty liver disease. Along with their wild-type (WT) littermates, heterozygous knockout mice for Gli2 (a proximal Shh signaling mediator) were fed a high-fat (HF) or a control diet for 12 weeks and their livers removed for gene expression profiling. All animals fed the HF diet showed an increase in body mass and liver steatosis, but Gli2+/- mice gained more weight and showed higher steatosis scores compared to the WT mice. Gli2+/- mice fed the control diet showed greater expression of genes involved in lipid metabolism, immune function, oxidative stress and cell cycle regulation, which suggests these mutants have an inherent metabolic defect impacting liver function. Top molecules include Cste, Taf1d, Gada45g, Sora7f4a, Lom2, Has3, Cyp2a5, and Hspa1a. These genes are involved in the Shh pathway genes did not show significant differences in gene expression. Though both Gli2+/- and WT mice fed the HF diet showed changes in genes involved in the same cellular processes, the Gli2+/- mice revealed a higher number of genes and genes associated with more advanced stages of liver disease. These results have key implications for the management of HPE patients and their families.

A simple abdominal ultrasound may help identify possible mutation carriers in families with history of HPE. Future studies will involve complete metabolic/endocrinological profiling and histological evaluation of the hypothalamic/pituitary axis of the Gli2+/- mice, and the study of additional animal models (currently, Shh+/- mice). Our ultimate goal is to identify genetic risk factors for idiopathic fatty liver in the general population, and genes associated with more advanced stages of liver disease. These results have key implications for the management of HPE patients and their families. A simple abdominal ultrasound may help identify possible mutation carriers in families with history of HPE. Future studies will involve complete metabolic/endocrinological profiling and histological evaluation of the hypothalamic/pituitary axis of the Gli2+/- mice, and the study of additional animal models (currently, Shh+/- mice). Our ultimate goal is to identify genetic risk factors for idiopathic fatty liver in the general population, and genes associated with more advanced stages of liver disease. These results have key implications for the management of HPE patients and their families.

2902F

The mouse is a widely used model organism well-suited for the study of mammalian development and physiology, and for elucidating details of disease mechanisms. Recent advances in imaging and genetic technologies have greatly expanded the options available to researchers to precisely label specific cellular subsets. The Jackson Laboratory (JAX) Mouse Repository, as part of its mission, facilitates the access of these valuable mouse strains to the research community. A variety of approaches used in mouse engineering mice incorporate both constitutive and inducible strategies, affording researchers a great deal of flexibility in controlling spatial and temporal expression. Numerous models employ a promoter with a well-characterized expression pattern, driving transcription of a reporter molecule. Among this growing set are GFP-expressing reporter lines developed by the Allen Institute for Brain Science (AIBS), various lacZ-expressing lines, and the latest generation of the well-known multi-colored ‘Brainbow’ models. The ‘Brainbow’ system is similar in concept to ‘Brainbow’ lines but with the difference being that they can be expressed in any tissue and cell type. A large number of strains that express cre in the brain are available, including those generated by the NIH Neuroscience Blueprint Cre Driver Network and the AIBS. The cre activity patterns for many of these strains have been characterized in a comprehensive manner at AIBS and JAX. The cre portal database (www.creportal.org) facilitates identification of an appropriate cre-expressing line by providing tools to search by promoter gene or by the anatomical site of expression. In combination with the wide array of cre reporter lines available, these resources enable the user to select cell subsets of choice with the preferred fluorescent or other reporter molecule by simply mating the appropriate cre-expressing and cre-reporter lines. These alleles can be used in conjunction with existing mouse models of human disease to aid in translational studies by providing a way to better image tissues involved in disease pathologies. Donating a strain to the JAX Mouse Repository is an easy way to fulfill the NIH’s requirements for sharing mice. Researchers wishing to have strains considered in the Repository are encouraged to submit their strains at: www.jax.org/donate-a-mouse.
2905F
Dysregulation of inflammatory pathways in a Familial Hemiplegic Migraine 1 mouse model after the induction of cortical spreading depression. E. Eising1, B. de Vries1, R. Styh1, L.S. Vrijhuijzen1, L.A.M. Broos1, N.A. Datson1, E.A. Tolner1, P.A.C. Hoien1, M.D. Ferrar1, A.M.J.M. van den Maagdenberg1, 1) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 2) Department of Neurology, Leiden University Medical Centre, Leiden, The Netherlands.

Familial Hemiplegic Migraine type 1 (FHM1) is a rare monogenic subtype of migraine with aura caused by mutations in the CACNA1A gene. CACNA1A encodes a subunit of voltage-gated CaV2.1 calcium channels that regulate neurotransmitter release in brain synapses. In FHM1 knock-in mouse models, these mutations enhance cortical glutamatergic signaling and increase the susceptibility for cortical spreading depression (CSD): a transient wave of neuronal and glial depolarization followed by a prolonged depression of excitatory neurotransmission. Reduction of cortical excitability was found to be caused by defects in the extracellular degradation of glycosylceramide. Individuals affected with hereditary spastic paraplegia 46 (SPG46) and autosomal-recessive cerebellar ataxia carried mutations in the β-glucosidase 2 gene (GBA2). Similar to glucocerebrosidase, GBA2 cleaves glucosylceramide, but GBA2 is located in the ER and/or at the plasma membrane. GBA2 activity was reported to be drastically reduced in a lymphoblastoid cell line derived from mononuclear white blood cells from an SPG46 patient. Assessment of enzyme activity of GBA2 is thus critically relevant. This raises the challenge to distinguish the GBA2 activity from that of glucocerebrosadase, as both β-glucosidases can degrade the same natural and artificial substrates under similar conditions. The first approach proposed to specifically measure GBA2 activity was to use conduritol B epoxide (CBE), which is a non-specific and irreversible inhibitor of β-glucosidase. CBE was found to inhibit not only GBA2, but also leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using N-butyld-glucosidase (NB-DGJ), which inhibits GBA2, but leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using N-butyld-glucosidase (NB-DGJ), which inhibits GBA2, but leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using N-butyld-glucosidase (NB-DGJ), which inhibits GBA2, but leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using N-butyld-glucosidase (NB-DGJ), which inhibits GBA2, but leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using N-butyld-glucosidase (NB-DGJ), which inhibits GBA2, but leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using N-butyld-glucosidase (NB-DGJ), which inhibits GBA2, but leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using N-butyld-glucosidase (NB-DGJ), which inhibits GBA2, but leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using N-butyld-glucosidase (NB-DGJ), which inhibits GBA2, but leaves glucocerebrosadase fully active. Therefore, we redefine GBA2 activity as the extent of GBA2 inhibition dependent on the CBE concentration and time of exposure.
Species-specific expression of acidic mammalian chitinase in stomach tissues. M. Ohno1, Y. Kogashi1, K. Tsuda2, K. Okawa3, M. Kamaya4, M. Sakaguchi1, Y. Sugahara1, F. Oyama1, 1) Biotechnology Laboratory, Dept Applied Chemistry, Kogakuin Univ; 2) Environment Analytical Chemistry Laboratory, Department of Environmental and Energy Chemistry, Kogakuin Univ. Two active chitinolytic enzymes, chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase), have been identified in mouse and human, although mammals do not have chitin synthase. Since these chitinolytic enzymes are significantly increased in Gaucher disease and mouse models of asthma, both chitinases have been regarded primarily as a host-defense mechanism against chitin-containing pathogens and/or play important roles in the pathophysiology. Little is known, however, about the mutual regulation of both chitinases between human and mouse. We quantified expression levels of the chitinases to compare the mRNA levels between human and mouse tissues on the same scale using a human-mouse hybrid standard DNA. Our analysis showed that Chit1 mRNA is expressed at similar levels in normal human and mouse lung. On the other hand, AMCase is predominantly overexpressed in mouse but not human stomach. To see whether these mRNA changes are reflecting actual differences in the levels of protein expression, we measured chitinolytic activity at pH 2.0 and pH 5.0 using the synthetic substrate of 4-methylumbelliferyl β-D-N, N-diacyctethylbiotiose (4MU-chitobiote) and detected robust chitinolytic activity in mouse stomach extract at pH 2.0 and strong activity at pH 5.0. In contrast, no activity was detected in that of human at pH 2.0 and very low activity was observed at pH 5.0. Furthermore, the anti-AMCase antibodies recognized a robust single protein band in extract from mouse stomach but not from human. Thus, mRNA differences between human and mouse stomach tissues were reflecting differences in both chitinase activity and levels of protein expression.


Chitinase hydrolyzes chitin, which is an N-acetyl-D-glucosamine polymer that is present in a wide range of organisms, including insects, parasites and fungi. Although mammals do not have chitin synthase, genes encoding chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase) and their translation products have been found in both human and mouse. Recent studies have shown an association between the mammalian chitinases and inflammatory diseases. For instance, Chit1 levels were elevated in the plasma of Gaucher disease and the bronchoalveolar lavage fluid of smokers with chronic obstructive pulmonary disease (COPD). AMCase expression and activity are upregulated in allergic airway responses in mouse models of asthma. Little is known, however, about mutual regulation of gene expression between Chit1 and AMCase in vivo. We established a quantitative PCR system using a single standard DNA and found that both chitinases mRNAs are predominantly expressed in mouse stomach. In stomach, AMCase mRNA is expressed 7-10 fold higher than those of housekeeping genes. These results indicate that the Chit1 and AMCase expression levels are relatively low in the human tissues examined and that the AMCase expression level in the stomach differs significantly between human and mouse.

Molecular basis of a new form of hyperekplexia. J. Capo-chichi1,2, S. Boiselle3, E. Brustein3, F.F. Hamden1,4, M. Samuels, G.A. Rouleau4, P. Drapeau4, J.J. Michaud2,1. 1) Centre de Recherche du CHU Sainte-Justine, Montréal, Québec, Canada; 2) Département de biochimie, Université de Montréal, Québec,Canada; 3) Département de pathologie et biologie cellulaire,Université de Montréal,Québec,Canada; 4) Centre de recherche du CHUM, Montréal,Québec,Canada.

Hyperekplexia (HK) is characterized by increased tone with startle reactions inducible upon tactile or auditory stimulations. HK is a genetic condition caused by disruption of glycine signaling in the CNS. Here we report on a consanguineous family from Cambodian origin composed of 4 affected female siblings who, in addition to severe HK, showed small birth weight, microcephaly with simplified gyral structure, and burst suppression on the EEG. They could not brea through themselves and died a few days after birth. We performed homozygosity mapping on all 4 affected siblings followed by exome sequencing on 2 of them and identified, in the shared regions of homogyzosity, 1 rare predicted-damaging homozygous mutation in CLPB (p.Ile562fs) that segregated with the disease. CLPB belongs to a subfamily of AAA+ proteins that includes clpb (bacteria) and hsp104/hsp78 (yeast), both of which function in protein disaggregation and resolubilization. Members of this subfamily form homo-oligomeric rings. p.Ile562fs abrogates CLPB function. To investigate the functional consequence of p.Ile562fs, we interfered with the zebrafish ortholog Clpb mRNA translation using a splice-blocking antisense morpholino (AMO). The specificity of the morpholino was confirmed by RT-PCR and western blot. The Clpb KD zebrafish larvae were well formed but presented some morphological and functional deficits (smaller eyes, shorter head and body length) compared to wild-type, non-injected larvae. They also exhibited a motor phenotype characterized by an abnormal touch-evoked response with increased maximum swim velocity and faster tail beat frequency. All larvae co-injected with wild-type human CLPB mRNA and Clpb AMO exhibited restoration of the morphological and motor phenotypes. Conversely, larvae co-injected with mutated human CLPB mRNA and Clpb AMO showed similar morphological and motor phenotypes to those observed in Clpb KD larvae. Thus, CLPB is a candidate gene for HK in this family.

A homozygous missense mutation in HSPA9 causes epiphysseal-vertebral-ear (EVE) dysplasia. J. Amiel1,2, M. Oufti3,4, A. Lingtard4, D. Lehallier5,6, C. Bole2, P. Nithake7, A. Munnoch1,2, S. Lyonnet1,4, A. Linnart1,2, P. Boissel1,2, P. Nithske1,4, F. F. Hamdan3,4, T. Boiselle3, E. Brustein3, F. F. Hamden1,4, M. Samuels, G. A. Rouleau4, P. Drapeau4, J. J. Michaud2,1. 1) Centre de Recherche du CHU Sainte-Justine, Montréal, Québec, Canada; 2) Département de biochimie, Université de Montréal, Québec, Canada; 3) Département de pathologie et biologie cellulaire, Université de Montréal, Québec, Canada; 4) Centre de recherche du CHUM, Montréal, Québec, Canada.

Eve and CODAS syndromes have not been reported. We performed exome sequencing of the two affected EVE dysplasia patients and after filtering for published and in-house SNPs and selecting for deleterious variants, the only deleterious missense variant shared between both sisters was the missense mutation p.Thr362Ile in heat shock 70kDa protein 9 (HSPA9). The catalytic activity of HSPA9 is encoded by the ATPase domain, a structure conserved in chaperones from vertebrates to bacteria. Thr362 is absolutely conserved in known orthologues of HSPA9, falls within the ATPase domain of AAA+ proteins that includes clpb (bacteria) and hsp104/hsp78 (yeast), of which function in protein disaggregation and resolubilization. Mem-
2913T
Autosomal Recessive Congenital Ichthyosis (ARC) and related disorders: Mutation in CERS3, coding for ceramide synthase 3, reveal major puzzle pieces for the understanding of epidermal barrier formation. K.M. Eckl1,2, R. Tidran3, H. Thiele4, V. Oji5, F. Hausser6, S. Brodesser1, M.L. Preiß1, A. Önal-Akan7, F. Stock8, D. Müller9, K. Becker10, R. Casper10, G. Nürnberg10, J. Altmüller10, F. Nürnberg6,8, H. Traupe6, A.H. Futerman5, H.C. Hennis1,2,6,8 1) Center for Dermatogenetics, Innbruck Medical University, Innsbruck, Austria; 2) Center for Dermatogenetics, Cologne Center for Genomics, University of Cologne, Cologne, Germany; 3) Dept. of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel; 4) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 5) Dept. of Dermatology, University Hospital of Münster, Münster, Germany; 6) Dept. of Dermatology, University Hospital of Heidelberg, Heidelberg, Germany; 7) Inst. of Med. Microbiology, Immunology and Hygiene, University of Cologne, Cologne, Germany; 8) Cluster of Excellence on Cellular Stress Responses in Aging-associated Diseases, University of Cologne, Cologne, Germany; 9) Practice for Dermatology Drs. Krnjaic, Merk, Preil, Schäfer, Ansbach, Germany; 10) Inst. of Human Genetics, University Hospital of Leipzig, Leipzig, Germany.
The barrier function of the human epidermis is supposed to be governed by lipid composition and organization in the stratum corneum. We report on a large German family with multiple consanguinity. The pedigree consists of several branches each with affected and unaffected family members. Using a combined approach of SNP-based autozygoscy mapping and exome sequencing in autosomal recessive congenital ichthyosis we identified a homozygous interval of 3.4 Mb on chromosome 15 containing homozygous missense mutations in two genes. Co-segregation analysis pinpointed mutations in CERS3, coding for ceramide synthase 3 (CerS3), as underlying the phenotype. Using LC-ESI-MS/MS we demonstrate that the mutation impairs the activity of CerS3, which synthesizes very long chain ceramides in the skin. We show a specific loss of free and bound ceramides with chain lengths from C26 up to C34 in terminally differentiating patient keratinocytes. Reconstructed skin from either healthy control cells or patient fibroblasts and keratinocytes reveals premature epidermal differentiation and a moderate impairment of barrier function in patient skin models. Our findings demonstrate that the chain length of ceramides is crucial early step for the formation of the epidermal barrier and indicate that disorders characterized by ichthyosis can be attributed to defects in the epidermal metabolism of ceramides and other components of the cornified envelope. We propose a link between several genes implicated in congenital ichthyosis has been identified, and we propose a crucial epidermal pathway of the lipid metabolism involving CERS3 and the ARCI genes TGM1, ALOX12B, ALOX5, ABCA12, CYP4F22, but also for LPIN, mutated in late onset autosomal ichthyosis; as well as for FALDH, ELOVL4, FATP4, ABHD5, which are mutated in syndromic ichthyosis disorders.

2915W
NPHP10 (SDCCAG8) interacts with components of the multi-aminocarboxyl-tRNA synthetase complex. K. Weihbrecht1,2,3, M. Humbert1,2,3, V. Sheffield1,2,4, S. Seo2 1) Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA; 2) Dept. of Ophthalmology and Visual Sciences, Univ. of Iowa, Iowa City, IA; 3) Howard Hughes Medical Institute, Univ. of Iowa, Iowa City, IA.
Nephronophisis (NPHP) is a recessive kidney disorder that is the leading cause of early onset, end-stage renal failure. Many patients mutated in cystic kidney disease have been shown to localize to the primary cilium and centrosomes, providing a coalescing mechanism for NPHP-related ciliopathies (NPHP-RC). Aside from renal failure and kidney cysts, retinal degeneration and dysplasia or degeneration of the cerebellum are also seen in many NPHP-RCs. SDCCAG8 is a nephronophisis gene (NPHP10), with patients exhibiting retinal and renal abnormalities, obesity, and learning disabilities. Mutations in SDCCAG8 were also found in several BBS patients, making SDCCAG8 the 16th BBS gene (BBS16). However, little is known about the molecular functions of NPHP10 and how loss of NPHP10 function leads to the observed phenotypes. Our goal was to gain insight into the function of NPHP10 by determining its interactors. Here, we show that NPHP10 interacts with components of the multi-aminocarboxyl-tRNA synthetase complex (MSC), including 8 out of 9 aminocarboxyl-tRNA synthetases (ARS) as well as aminoacyl-tRNA synthetase-complex interacting multifunctional protein 2 (AIMP2). We performed tandem affinity purification using stably transfected HEK293T cells expressing FLAG- and S-tagged NPHP10 and isolated its associated proteins. We further determined that among the MSC components, NPHP10 directly interacts with AIMP2. Co-immunoprecipitation was used to determine interacting domains of these two proteins. Sucre gradient ultracentrifugation of WT mouse tissues showed that NPHP10 is associated with the MSC in the kidney, brain, and eye, all affected tissues in NPHP10 patients. Lastly, we show that disruption of the MSC through AIMP2 knockdown significantly reduces ciliogenesis. Altogether, these findings suggest that NPHP10 functions through an MSC-mediated mechanism or vice versa, and that the MSC is an important component for ciliary function.

2916T
Fragile X Mental Retardation Protein (FMRP) in cell differentiation: The MEG-01 as a new study model. M. McCoy, F. Corbin, Biochemistry Dept, University of Sherbrooke, Sherbrooke, Quebec, Canada.
Introduction: Fragile X Syndrome (FXS) is the most common form of inherited mental retardation. It is triggered by the expansion of CGG repeats in the 5'UTR of the FMR1 gene, as well as from the methylation of its CpG island. The latter causes transcriptional silencing of the encoded Fragile X Mental Retardation Protein (FMRP). FMRP is an RNA-binding protein commonly associated with polyribosomes and is thought to play a critical role in the maturation of the synaptic network by repressing translation. Consequently, in the absence of FMRP, deregulated protein synthesis results in the abnormal maturation of neuronal dendritic spines. To investigate the more universal role of FMRP in architecture development during cell differentiation, a new model was established using a megakaryoblastic cell line: MEG-01. Cell fractionation experiments, confocal microscopy and Western Blot analysis were used to determine interacting domains of these two proteins. Complementary techniques were employed to study the behavior of FMRP during those changes. In fractionation experiments, proteins were collected at different stages of cell differentiation, either by differential ultracentrifugation or on sucrose gradient. Protein content was then revealed by Western Blot analysis. Inimmunofluorescence experiments, protein expression was analyzed by FACS, while confocal microscopy was performed for subcellular localization and colocalization assays. Results: Cell fractionation experiments revealed the redistribution of FMRP and other RNA-binding proteins throughout the cell upon activation. By microscopy, the reorganization of the cytoskeleton and the elongation of filopodial extensions, typical of platelet-like particles (PLPs), was observed. FMRP and other RNA-binding proteins were shown to colocalize with microtubules in the cytoplasmic extensions. Conclusion: The regulation of local synthesis appears to be important in the maturation of the synaptic extensions, considering that FMRP and other proteins of the translational machinery accompany mRNPs towards their sites of translation. The reorganization of the cytoskeleton and colocalization of mRNPs with microtubules suggests that translational complexes are carried from the core of the cell to filopodial extensions for local synthesis by means of microtubular transport, as is the case in dendritic spines of neurons. The FMRP-FMRP binding complex could be a key factor for regulating the process of maturation of elaborate cellular networks.
2917F
Hypervitaminosis D due to 1,25-(OH)2D-24 hydroxylase (CYP24A1) deficiency causing nephrocalcinosis and nephrolithiasis. G. Nesterova1, M. Malidcan2, T. Sakaki1, M. Collins3, D. Adams4,5, C. Boerkoel5, W. Gahl2,6, 1) Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, USA; 2) Department of Biotechnology, Faculty of Engineering, Toyama Prefectural University, Toyama, Japan; 3) Skeletal Clinical Studies Unit, Craniofacial and Skeletal Diseases Branch, National Institute of Dental and Craniofacial Research, NIH, Bethesda, USA; 4) NIH Undiagnosed Diseases Program, NIH Office of Rare Diseases Research and Human Genome Research Institute, Bethesda, MD, USA; 5) Office of the Clinical Director, National Human Genome Research Institute, NIH, Bethesda, MD, USA.

Background/Objectives: An excess of an active form of Vitamin D1, or 1α,25(OH)2D3, can result in nephrocalcinosis and nephrolithiasis. Nephrolithiasis represents a global health problem with prevalence at 10%; we evaluated the cause of increased 1α,25(OH)2D3 levels in the development of those disorders. Patients and Methods: We measured serum and urine calcium and phosphate, vitamin D metabolites and performed mutation analysis in CYP24A1 gene encoding 1,25(OH)2D-24-hydroxylase in two patients with hypervitaminosis D, nephrocalcinosis or nephrolithiasis enrolled in the National Institutes of Health Undiagnosed Diseases Program. CYP24A1 pathologival variants and mutations were evaluated using the dbSNP database and the pathogenicity prediction programs. Results: Both patients exhibited hypercalciuria, low level of 24,25(OH)2D3, elevated 1α,25(OH)2D3, and undetectable activity of 1,25(OH)2D-24-hydroxylase that inactivates 1α,25(OH)2D3. They had biallelic mutations in the CYP24A1 leading to the loss of function of this enzyme. Based upon dbSNP data, the frequency of deleterious biallelic CYP24A1 mutations and variants is estimated to be 1960 per 100,000 in the general population. That would be a significant number contributing to the prevalence of nephrocalcinosis.

Conclusion/Discussion: We found that CYP24A1 loss of the function mutations is associated with nephrocalcinosis and nephrolithiasis. Our assessment of CYP24A1 gene variants predicted that pathogenic defects in CYP24A1 may account for a significant fraction of nephrocalcinosis. Recognition of CYP24A1 deficiency could prompt the recommendation that 1α,25(OH)2D3 levels be determined in such patients and promotes therapy for hypervitaminosis D.

2918W
ARID1B inhibits WNT signaling through interaction with BRG1 and β-catenin. G. Vasileiou1, A.B. Etki1, S. Jebre1, J. Behrens2, A. Reis1, M.V. Hadjihannas3, 1) Inst. of Human Genetics, Universität Erlangen-Nürnberg, Erlangen, Germany; 2) Nikolaus Fiebiger Center, Universität Erlangen-Nürnberg, Erlangen, Germany; 3) Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD, USA.

The SWI/SNF chromatin remodeling complex regulates transcription by remodeling chromatin and through interactions with transcriptional regulators. Recently, deleterious mutations in subunits of this complex, which carry enzymatic activity and confer functional specificity, have been found in a diverse spectrum of human cancers, with an overall frequency approaching that of mutations in the p53 oncogene. Likewise, haploinsufficiency due to mutations in various components of the SWI/SNF complex was shown to be a frequent cause of the Coffin-Siris syndrome and intellectual disability. The mechanisms by which mutations in SWI/SNF subunits contribute to cancer and intellectual disability remain poorly understood. Here we show that ARID1B suppresses Wnt/β-catenin signaling and that Wnt/β-catenin target genes are upregulated in intellectual disability patients carrying mutations in ARID1B as compared to normal subjects. We found that ARID1B associates with β-catenin in nuclear puncta through a mechanism involving the SWI/SNF subunit BRG1, a binding partner of both β-catenin and ARID1B. Mutation of ARID1B as seen in an intellectual disability patient leads to partial truncation of its BRG1 binding domain, which reduces the ability of ARID1B to localize β-catenin to nuclear puncta and interferes with the suppressive effect on β-catenin-driven transcription. Aberrant activation of Wnt/β-catenin signaling is a causative feature of many human cancers and has been previously linked to brain development. Our results offer an explanation as to how the tumor suppressive effect of chromatin remodeling complexes is mediated and specifically how mutations in ARID1B lead to some of the clinicopathological features observed in syndromic and non syndromic intellectual disability. Finally, our results indicate that while chromatin remodeling can have both activating and repressive effect on transcription, its repressive function could play an important role in human development and cancer through modulation of developmental/oncogenic pathways.

2919T
Cardiovascular manifestations in a family with GLA nonsense mutation (W162X), E. Severin1, C. Dragomir2, A. Stan2, G. Sarca2, 1) Depat Human Genetics, Carol Davila Univ Med Pharm, Bucharest, Romania; 2) Genetic Lab, Bucharest, Romania.

Background: Previous studies reported that different mutations of GLA gene are associated with multigorgan damage. A three-generation Caucasian family with alpha-galactosidase A gene mutation is described. Complete or partial deficiency of alpha-galactosidase A results in the accumulation of globotriaosylceramide in cells throughout the body, leading to severe and progressive cardiovascular manifestations. Despite being recessive X-linked, carrier females are affected by Fabry disease and can express the same cardiovascular symptoms as males. Objectives: To describe and compare cardiovascular findings, to confirm the diagnosis of Fabry disease in male and female family members and to identify the mutation causing variable symptoms of cardiovascular phenotype. Patients and Methods: 2 males and 5 females of the family were enrolled and clinically assessed (including echocardiography); enzyme activity levels were also evaluated. Genomic DNA was isolated from blood samples of all family members and analyzed for GLA gene mutation. Results: Both male patients were found hemizygotes having one copy of GLA gene mutation (c.485G>A) and no enzyme alpha GAL activity. Also, all five females were found carrier of the same mutation and a broader range of residual alpha GAL activity. Despite having the same mutation, they exhibited different cardiovascular phenotypes. One male patient (31-year-old) had no cardiac involvement but his maternal uncle (56-year-old) presented left ventricular hypertrophy, cardiomyopathy and cerebrovascular manifestations. Mild left ventricular hypertrophy and moderate interventricular septum thickening were noticed in females by the third decade of life. One female was found normal by echocardiography; enzyme activity levels were also evaluated. Genomic DNA was isolated from blood samples of all family members and analyzed for GLA gene mutation. Results: Both male patients were found hemizygotes having one copy of GLA gene mutation (c.485G>A) and no enzyme alpha GAL activity. Early genetic testing should be considered in younger female with a positive family history of Fabry disease. ERT should be instituted at an earlier age, particularly in carrier females with low alpha GAL activity, in order to limit severe expression of Fabry disease.
2921W


Background: Mutations in transcription factors involved in pituitary organogenesis and in the GHRH-GH axis genes are involved in the etiology of congenital GH deficiency. Objective: to diagnose the etiology of congenital GH deficiency by selecting candidate genes based on a) the deficiency of GH alone (isolated GH deficiency, IGHD) or associated to other pituitary hormones (combined pituitary hormone deficiency, CPHD), and b) on magnetic resonance imaging (MRI) studies of the pituitary gland. Patients and Methods: Retrospective analysis of genetic results in 207 patients with congenital GH deficiency. GH deficiency was diagnosed by failure to respond to two GH-stimulation tests: clonidine test followed by combined insulin, TRH and GnRH test, which also evaluated other pituitary hormone deficiencies. GH1, GHRHR and GHRHR were studied only in patients with IGHD. PROP1 was studied in CPHD patients with topic posterior pituitary lobe (PPL). HESX1 and GLI2 were studied in patients with IGHD and CPHD irrespective of MRI findings. DNA was extracted from peripheral leukocytes and the coding regions and intronic boundaries of the respective genes were amplified by PCR and sequenced by Sanger method. Results: Among 57 patients with IGHD, 8 had homozygous GH1 mutations, 7 (5 families) homozygous GHRHR mutations, all with a PPL in the normal position, but none GH1 mutations. Among the 140 patients with IGHD, 11 (9 families) had homozygous PROP1 mutations, 1 patient a homozygous HESX1 mutation and 27 patients heterozygous GLI2 mutations (3 frameshift/stop codon and 24 non-synonymous; 26 CPHD and 18 ectopic PPL). One GLI2 mutation was associated to GH deficiency in 4 additional family members. In one family GLI2 mutation was associated to polydactyly and in another to midline facial defects. Holoprosencephaly and septo-optic dysplasia were not present. The pattern of inheritance was autosomal recessive and segregation was observed in all family members. Conclusions: Mutations in transcription factors involved in pituitary organogenesis and in the GHRH-GH axis genes are involved in the etiology of congenital GH deficiency. Hormonal testing and MRI of the pituitary are explained the etiology of congenital GH deficiency in 13% of patients. Heterozygous GH1, GHRHR and autosomal dominant with incomplete penetrance and variable expression. The recurrence of mutations affecting these highly conserved regions involved in interaction with the transported cations in two paralogous subfamilies is suggestive of a gain-of-function effect. Functional in vitro studies of ATP1A1 mutants provided evidence for inappropriate depolarization of cells with ATPase alterations, suggesting cation influx. In summary, gain-of-function alterations in two members of the ATPase family result in autonomic adrenomedullary secretion in roughly 7% of aldosterone-producing adenomas. The same mutational mechanism might be causative for other endocrine tumors. We will present our investigations in further endocrine tumor entitles.

2922T

Somatic mutations in ATP1A1 and ATP2B3 lead to aldosterone-producing adenomas and secondary hypertension. T. Wieland1, P. Beuschelein2, S. Bruse3, J. Renner4, H.N. Nielsen5, U.D. Leonard6, V.R. Schack6,5, L. Amar1,4, E. Fischer2, A. Walther2,1, P. Tauber2,1, T. Schwarzmayr1, S. Diener7, E. Graf7, B. Alloio8, B. Sanmson-Coutière1,4, A. Benecke9, M. Quinkler10, F. Fallo11, P.F. Plouin1,2, F. Mantero1,2, T. Mägdefrau1,2, P. Mutikainen1,3, M. Vanasse1,9, M. Zennaro1,13, P.M. Mulatero1,14, C. Zennaro1,13,14, P. Tauber1,2,15, V.R. Schack1,2, L. Amar1,4, E. Graf7, B. Alloio8, B. Sanmson-Coutière1,4, A. Benecke9, M. Quinkler10, F. Fallo11, P.F. Plouin1,2, F. Mantero1,2, T. Mägdefrau1,2, P. Mutikainen1,3, M. Vanasse1,9, M. Zennaro1,13, P.M. Mulatero1,14, C. Zennaro1,13,14, T.M. Strom1,15, M. Reincke2,1.

1) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 2) Medizinische Klinik und Poliklinik IV, Ludwig-Maximilians-Universität München, Munich, Germany; 29) Institut National de la Santé et des Recherches Médicales (INSERM), Unité Mixte de Recherche Scientifique (UMR) 970, Paris Cardiovascular Research Center, Paris, France; 4) Université Paris Descartes, Sorbonne Paris Cité, Paris, France; 5) Department of Biomedicine, Aarhus University, Aarhus, Denmark, Danois National University of Science and Technology; 6) University Regensburg, Regensburg, Germany; 7) Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Paris, France; 8) Department of Medicine I, Endocrine and Diabetes Unit, University Hospital Würzburg, Würzburg, Germany; 9) Centre National de la Recherche Scientifique (CNRS), Institut des Hautes Etudes Scientifiques, Bures sur Yvette, France; 10) Clinical Endocrinology, Campus Mitte, University Hospital Charité, Berlin, Germany; 11) Department of Medicine, University of Padova, Padova, Italy; 12) Endocrine Unit, Department of Medicine, University of Padova, Padova, Italy; 13) Institute of Human Genetics, Technische Universität München, Munich, Germany; 14) DZHKG (German Centre for Cardiovascular Research), partner site, Munich Heart Alliance, Munich, Germany; 15) Department of Medical Sciences, Division of Internal Medicine and Hypertension, University of Turin, Turin, Italy; 16) Department of Medical Sciences, Division of Internal Medicine and Hypertension, University of Padova, Padova, Italy; 17) Centre National de la Recherche Scientifique (CNRS), Institut des Hautes Etudes Scientifiques, Bures sur Yvette, France; 18) Endocrine Unit, Department of Medicine, University of Padova, Padova, Italy; 19) Center for Integrative Biology, University of Trento, Trento, Italy; 20) Institute of Human Genetics, National Hospital of Pediatrics, Hanoi, Hanoi, Viet Nam; 21) 1) Human Genetics, National Hospital of Pediatrics, Hanoi, Hanoi, Viet Nam; 2) Institute of Genome Research, Hanoi, Vietnam.


1) Human Genetics, National Hospital of Pediatrics, Hanoi, Hanoi, Viet Nam; 2) Institute of Genome Research, Hanoi, Vietnam.

Con genital adrenal hyperplasia (CAH) is an autosomal recessive disease which is characterized by a deficiency of one of the enzymes involved in the synthesis of cortisol from cholesterol by the adrenal cortex. More than 90% of CAH cases are due to 21-hydroxylase deficiency (21OH, whereas 5-8% arise from 11β-hydroxylase deficiency. Mutations in the CYP11B1 gene cause the gene splicing during transcription. This CYP11B1 genotype testing is the basis for genetic counseling and prenatal diagnosis in the future.
2924W
Alterations in the PAX8 promoter region cause thyroid dysgenesis. P. Hermanns1, M. Mortel2, M. Donaldson3, J. Jones1; P. Pohlenz2. 1) Molecular Endocrinology, University Medical Center Mann, Mainz, Germany; 2) Endokrinologikum Hannover, Centre for Hormone and Metabolic Diseases, Hannover, Germany; 3) 3Department of Child Health, Royal Hospital for Sick Children, Glasgow, Ireland.
Thyroid dysgenesis (TD) is the cause of approximately 80% of patients diagnosed to have congenital hypothyroidism. So far, mutations in five candidate genes (PAX8, TSHR, TTF1, TTF2 and NKX2.5, respectively) have been identified to cause TD. We screened 190 patients with TD for mutations in these genes. Besides previously described and characterized mutations, we detected new so far unreported mutations in the coding regions of PAX8, TSHR, TTF1, TTF2 and NKX2.5. 100 normal individuals did not harbour those mutations. Interestingly, in addition to mutations in the coding region of the known candidate genes we also detected four different base pair exchanges in the PAX8 promoter region. Very recently, we have found one of these base pair changes in a patient diagnosed with thyroid dysgenesis who was heterozygous for a mutation in the NKX2.5 gene and heterozygous for a base pair change in the PAX8 promoter. In vitro studies were performed to unravel the mechanisms by which these newly identified promoter base pair exchanges might be causative. Electromobility shift assay (EMSA) studies suggest no specific protein or protein complex binding to the altered promoter elements. Transient transfection studies in different cell lines showed that at least one of these base pair changes leads to a significantly decreased promoter activity and thus to an impaired PAX8 gene expression. In summary, we identified a new group of PAX8 promoter sequence alterations that might cause TD. Further studies are needed to prove this hypothesis.

2925T
Shortening the diagnostic odyssey of patients with very early onset inflammatory bowel disease. S. Drury1, J. Kammernier4, L. Jenkins1, M. Elawad5, K. Gilmore4, N. Leach6, N. Shah4. 1) NE Thames Regional Genetics Service, Great Ormond Street Hospital, London, London, United Kingdom; 2) Department of Gastroenterology, Great Ormond Street Hospital for Children, London, WC1N 3JH; 3) Centre for Translational Genomics-GGSGene, UCL Institute of Child Health, London, WC1N 1EH, United Kingdom; 4) Department of Immunology, Great Ormond Street Hospital for Children, London, WC1N 3JH.
Very early onset inflammatory bowel disease (VEO-IBD) results in chronic intractable and therapy resistant inflammation of the gut and severe growth failure. The pervasive nature of the disorder often requires lengthy hospital admission. Despite underlying genetic heterogeneity, clinical presentation is often homogeneous. The differential diagnosis of VEO-IBD includes mono- genic conditions with defects in epithelial barrier function, T/B lymphocyte selection and activation and innate and adaptive immune response as well as conditions leading to neutropenia and defective phagocyte killing and hyper- and autoinflammation. We have implemented a next generation sequencing approach to expedite the diagnosis of VEO-IBD patients, by screening 40 genes associated with the disorder. Coverage of the panel overall is 99% at 30x or greater. Thus far 5 positive controls and 6 VEO-IBD patients have been screened. All control mutations were detected and the pathogenic mutation in 1 previously undiagnosed patient has been identified. The subject presented with severe intractable inflammation of the large and small bowel and was diagnosed with autoimmune enterocolitis. Despite lack of diagnosis and due to the severity of the condition the subject has undergone experimental haematopoietic stem cell transplant (HCST), which has a mortality rate of 20%. Subsequent molecular analysis, showed the patient to be a compound heterozygote for mutations in TFC37 (c.2018G>A p.Gly673Asp and c.2808G>A p.Trp936*), associated with tricho-hepato-enteric syndrome (THES), a very rare disease (1/1,000,000 births). The phenotype of this patient is unlike the characteristic features reported in other cases; thus NGS has enabled a diagnosis of a) a rare disorder not characterised by phenotype alone and b) a genetic test which until now has not been offered clinically. Due to the rarity of THES, it is currently unknown whether HCST is optimal treatment for the condition. Identifying the molecular basis of VEO-IBD has the potential to profoundly affect patient management; e.g. patients with mutations in IL10 can be successfully treated with HCST. In addition 42 primary immune deficiency genes are also captured with the same panel and analysis can be extended to these genes, offering a comprehensive molecular approach to this complex group of patients. We aim to introduce this extensive genetic screening early on in patient management, to guide medical intervention at the earliest opportunity.

2926F
Patient derived somatic and induced pluripotent stem cells as a model for functional assessment of mutations identified by exome sequencing in congenital diarrheal disorders. M. Youshav1, A. Vega-Crespo2, R. Solorzano-Vargas2, J. Wang2, S. Stanford2, C. Sosa3, S. Nelson4, J. Byrne5, M. Martin2. 1) Human Genetics, UCLA Geffen School of Medicine, Los Angeles, CA; 2) UCLA Geffen School of Medicine, Los Angeles CA 90024.
Background: Children presenting with intractable diarrhea often have a poor prognosis; after a sometimes fruitless diagnostic odyssey they can end up with prolonged intravenous nutrition and/or intestinal transplant. Morbidity and mortality are high, and medical care costs overwhelming. Our pilot study of the use of whole exome sequencing to diagnose difficult cases of congenital diarrhea found convincing genetic evidence of mutations in genes that are known to cause malabsorptive diarrhea or other conditions, including ADAM17 (inflammatory skin and malabsorptive diarrhea), EPCAM (Tufting enteropathy), MYO5B (diarrhea with microvillus atrophy), NEUROG3 (generalized malabsorptive diarrhea), PCSK1 (PC1/3 deficiency), SI (sucrase isomaltase deficiency), and SLCSA1 (glucose/galactose malabsorption), as well as a number of novel candidate genes. It is now possible to model these disorders in gut epithelium generated from somatic intestinal stem cells and induced pluripotent stem cells (iPSCs). Methods: We performed exome sequencing on cases with congenital malabsorptive, secretory, or IBD-like diarrhea having no convincing diagnosis after a conventional medical and genetic workup. To model these disorders we took two approaches: 1) isolate somatic intestinal stem cells from patients’ endoscopic biopsies or surgical samples from normal controls, or 2) derive induced pluripotent stem cells (iPSCs) from dermal fibroblasts obtained by skin biopsies, and subsequently inject the iPSCs towards intestinal epithelial fate. We used methods including RNAseq, immunofluorescence, and other assays to assess the functional effect of mutations identified by exome sequencing in patient cell lines, and normal cell lines in which selective genes are depleted using shRNA methods. Results: We developed disease-specific somatic intestinal stem cell and pluripotent cells lines, characterized them for reprogramming by assessing their expression of pluripotency markers, their karyotype, and their ability to form representatives of all three germ layers. We injected different iPSC lines into intestinal epithelium for disease modeling purposes and identified phenotypes caused by mutations in a number of the genes. Conclusion: The development of somatic intestinal stem cell and iPSC lines has allowed us to model severe diarrheal disorders within a dish and provides a potential foundation for future personalized cellular therapeutics for digestive system disorders.
2927W
Exome sequencing of a familial trio with a suspected autosomal dominant idiopathic immune deficient syndrome identifies novel candidate mutations in the complement system and two other genes. R. Gohari\textsuperscript{1}, J. Liang\textsuperscript{2}, D. Li\textsuperscript{1}, C. Kao\textsuperscript{1}, Y. Guo\textsuperscript{1}, W. Chen\textsuperscript{1}, L. Tian\textsuperscript{1}, F. Wang\textsuperscript{1}, J. Synder\textsuperscript{1}, N. Abdel-Magid\textsuperscript{1}, L. Vazquez\textsuperscript{1}, B. Keating\textsuperscript{1,3,4}, J. Zhang\textsuperscript{2}, H. Hakonarson\textsuperscript{1,3,4}, 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 4) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

We sequenced the exomes of a familial trio with an affected mother and child showing signs of an autosomal dominant idiopathic immune deficient syndrome to determine the potential underlying genetic cause. The affected child suffers from chronic sinusitis, conjunctivitis, constipation, and recurrent skin lesions. The mother suffers from chronic respiratory infections, recurrent bacterial and viral infections. Both probands have a history of bacteremia and maintain a low white-blood cell count during infections without fever. We suspected an autosomal dominant inherited deleterious mutation in one of the Toll-like receptors or within the TLR signaling pathway. The Toll-like receptors detect pathogens by recognizing specific microbial components. Exome sequencing revealed a splice acceptor site mutation in NEMO at chrX:153,792,177 (T→C). The child appeared homozygote, while the mother and father both appeared heterozygote. Knowing the father cannot be heterozygote, we determined the second half of NEMO is duplicated approximately 100kb downstream on the 3' strand. Follow-up Sanger sequencing of NEMO and the NEMO-homolog using long-range PCR show the child has the mutation in both NEMO and the NEMO-homolog, while the father has the mutation in NEMO but not in the NEMO-homolog. As the father is unaffected, this rules out NEMO as a candidate. We performed a more exhaustive variant reduction analysis revealing a novel heterozygote deletion in C2 (chr6:31902193-4, AG) and 2 heterozygote mutations, in LILRB1 (chr19:55143963, A/G) and KIR3DL3 (chr19:5256741, A/T). C2 is part of the classical complement system which interacts with the adaptive immune system to clear pathogens. We suspect the adaptive immune system is attempting to recruit the complement system, however this 2bp deletion may be recruiting a non-functional LILRB1. KIR3DL3 is the leucocyte immunoglobulin-like receptor family, and is expressed on immune cells. LILRB1 controls inflammatory responses to focus immune response and limit autoactivity. KIR3DL3 is killer cell immuno globulin-like receptor expressed by natural killer cells, subsets of T cells and is thought to play a role in regulation of immune response. Both LILRB1 and KIR3DL3 exist in highly variable regions and the genetic variation observed may be normal. All three mutations need further functional validation before we can conclusively attribute them to this idiopathic immune deficient syndrome.

2928T
SNPs associated with cerebrovascular accident in a Brazilian cohort of sickle cell anemia patients. P.R.S. Cruz\textsuperscript{1}, G. Ananina\textsuperscript{1,2}, F. Menas\textsuperscript{1,2}, M.R.C. Bozerra\textsuperscript{1}, A.S. Araujo\textsuperscript{1}, G.P. Gil\textsuperscript{2}, W.M. Avelar\textsuperscript{2}, F. Conde\textsuperscript{3}, F.F. Costa\textsuperscript{2}, M.B. Melo\textsuperscript{1,2}. 1) Center of Molecular Biology and Genetic Engineering (CBMEG), University of Campinas, Campinas, São Paulo, Brazil; 2) Hematology and Hemotherapy Center/HEMOCENTRO, University of Campinas, Campinas, São Paulo, Brazil; 3) Hematology and Hemotherapy Center of Pernambuco (HEMOPE), Recife, Pernambuco, Brazil; 4) Neuroimaging Laboratory, Department of Neurology, UNICAMP, Campinas, São Paulo, Brazil.

We suspected an autosomal dominant inherited deleterious mutation in one of the Toll-like receptors or within the TLR signaling pathway. The Toll-like receptors detect pathogens by recognizing specific microbial components. Exome sequencing revealed a splice acceptor site mutation in NEMO at chrX:153,792,177 (T→C). The child appeared homozygote, while the mother and father both appeared heterozygote. Knowing the father cannot be heterozygote, we determined the second half of NEMO is duplicated approximately 100kb downstream on the 3' strand. Follow-up Sanger sequencing of NEMO and the NEMO-homolog using long-range PCR show the child has the mutation in both NEMO and the NEMO-homolog, while the father has the mutation in NEMO but not in the NEMO-homolog. As the father is unaffected, this rules out NEMO as a candidate. We performed a more exhaustive variant reduction analysis revealing a novel heterozygote deletion in C2 (chr6:31902193-4, AG) and 2 heterozygote mutations, in LILRB1 (chr19:55143963, A/G) and KIR3DL3 (chr19:5256741, A/T). C2 is part of the classical complement system which interacts with the adaptive immune system to clear pathogens. We suspect the adaptive immune system is attempting to recruit the complement system, however this 2bp deletion may be recruiting a non-functional LILRB1. KIR3DL3 is the leucocyte immunoglobulin-like receptor family, and is expressed on immune cells. LILRB1 controls inflammatory responses to focus immune response and limit autoactivity. KIR3DL3 is killer cell immunoglobulin-like receptor expressed by natural killer cells, subsets of T cells and is thought to play a role in regulation of immune response. Both LILRB1 and KIR3DL3 exist in highly variable regions and the genetic variation observed may be normal. All three mutations need further functional validation before we can conclusively attribute them to this idiopathic immune deficient syndrome.

2929F
Clinical, Immunological, and Molecular Characterization of JAK3 Deficiency causing Severe Combined Immunodeficiency Disease in Saudi Arabia, A. Hawwari\textsuperscript{1}, H. Al-Shammari\textsuperscript{2}, S. Al-Hasi\textsuperscript{1}, O. A-Smadi\textsuperscript{1}, H. Al-Dheiri\textsuperscript{1}, A. Al-Ghonaum\textsuperscript{2}, S. Al-Muhse\textsuperscript{2}, B. AL-Saud\textsuperscript{2}, R. Arnaout\textsuperscript{1}, H. Al-Mousa\textsuperscript{2,3}. 1) Genetics, Research Center, MBC: 03, King Faisal Specialist Hospital & Research Center, PO Box 3354, Riyadh 11211, Saudi Arabia.; 2) Department of Pediatrics, King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia; 3) Al-Faisal University, Riyadh, Saudi Arabia.

Molecular defects in IL2RG are the most common genetic cause of T-B+NK- SCID worldwide followed by JAK3. However, no report has been published to describe the genetic defects of T-B+NK- SCID in Saudi Arabia. Due to high consanguinity rate in Saudi society, most cases that we observe in our patient population are mainly due to autosomal recessive pattern of inheritance. Hence, patients with T-B+NK- SCID were screened for mutations in JAK3 gene. Eight patients from 5 families diagnosed with T-B+NK- were screened for mutations in JAK3 gene. The diagnosis of T-B+NK- SCID is based on the profound, characteristic lymphopenia and early clinical presentation of immunodeficiency often with strong family history indicating an autosomal recessive inheritance. Five novel homologous JAK3 mutations were identified. Three closely related patients (two siblings and a cousin) carry the nonsense mutation (Q305X); one patient was found to have a deletion mutation (A3838568X14) leading to a stop codon 14 amino acid downstream; and two brother patients carry the R403C missense mutation. The remaining two patients carry either E1019K or R103H missense mutations. Contrary to the available data of X-linked IL-2 common gamma chain gene defect of T-B+NK- SCID as the most common, we demonstrate here that in the Kingdom of Saudi Arabia the most common form of T-B-NK- SCID is the autosomal recessive mode of inheritance associated with JAK3 deficiency.
2930W
Likely pathogenic hypomorphic mutation in the perforin 1 gene causing adult-onset familial hemophagocytic lymphohistiocytosis. L. Massingham1, J. Walsh1, N. Shurt2, C. Benson3, P. Rintela2, N. Berliner2, D. Treaba1, J. L4, C. Pompoluhuki1. 1) Department of Pediatrics, Division of Genetics, Hasbro Children's Hospital and Rhode Island Hospital, Providence, RI; 2) Division of Hematology, Rhode Island Hospital, Providence, RI; 3) Division of Hematology, Brigham and Women's Hospital, Boston, MA; 4) Department of Pathology, Rhode Island Hospital, Providence, RI.
Familial hemophagocytic lymphohistiocytosis (FHL) is characterized by fever and hepatosplenomegaly with cytopenias and hyperactivated macrophages and T-lymphocytes. FHL is inherited in an autosomal recessive manner and currently there are five known molecular subtypes (FHL-1-5) with overlapping phenotypes. The age of onset is generally in infancy or the first few years of life, but more recently adult-onset FHL has been reported. Distinguishing between FHL and secondary (acquired) hemophagocytic lymphohistiocytosis (HLH) is difficult and treatment recommendations vary considerably. Mutations in the perforin 1 (PRF1) gene account for 80% of patients with FHL. Case: A 30 year old previously healthy man presented with a 2 week history of cough and fever. Initial labs revealed p.Ala91Val variant may be hypomorphic. Our patient's cytopenia was consistent with disease caused by PRF1 mutations. IHC staining revealed splenomegaly and a diffuse pulmonary interstitial infiltrate. He was presumed to have hemophagocytic lymphohistiocytosis and chemotherapy was initiated. A persistently abnormal coagulation profile prompted a repeat bone marrow biopsy, which showed a hypocellular bone marrow with BM eosinophils and T-lymphocytes. PRF1 was confirmed by sequencing of the BM. FHL diagnosis was supported by the cytopenia and genetic counseling recommendation. Our team made the clinical diagnosis of FHL. "Our patient’s disease was diagnosed after a delay of several years, as no family history of FHL was evident. Our case highlights the importance of genetic testing and counseling for patients with atypical HLH."

2931T
Ancestry and admixture admixture among sickle cell disease patients in North America. Z. Wang1, L. Diaw1, M. Barr1, M. Quinn1, D. Diggis1, A. Ogahoe1, D. Darbari1, A. Hutchinson2, C. Hoppe3, J.G. Taylor4. 1) Genetics Medicine Section, Hematology Branch, NHLBI, NIH, Bethesda, MD, USA; 2) Center for Cancer and Blood Diseases, Children’s National Medical Center, Washington, DC, USA; 3) Cancer Genetics Research Laboratory, FAC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA; 4) Oakland Children’s Hospital, Oakland, CA, USA.
Genome wide association studies (GWAS) are an important tool for identifying complex human diseases. An important variable for GWAS is the degree of admixture occurring within study populations. If undetected, admixture can lead to spurious associations, although it may also be exploited as a strategy for mapping by admixture linkage disequilibrium (MALD). As part of our efforts to identify genetic modifiers of sickle cell disease (SCD), we have examined ancestry and genetic admixture in SCD patients living in North America. We first examined ancestry from family histories for 649 adults with SCD recruited to the Bethesda Sickle Cell Cohort Study in the eastern US, where 61% of subjects are African American, 22% African, 12% Caribbean or South American and 1% of others, suggesting a high degree of population diversity. We further examined these observations with genetic markers by identifying 364 ancestry informative markers (AIMs) from over 4 million SNPs typed in Europeans (CEU) and Yorubans (YRI) available from the Human HaploType Map. AIMs were typed in 445 of these SCD cohort subjects. Principle components analysis showed African American or Caribbean/South American subjects had the highest degree and a wider range of admixture compared to Africans with SCD or Hispanic populations (CEU, YRI and CHB). STRUCTURE analysis showed similar results. We also performed a replication study using 471 anonymous DNA samples from SCD subjects in a western US newborn screen cohort. Similar to the analysis of population ancestry in the Bethesda cohort, the SCD newborns are highly admixed, plotting between YRI and CEU populations. However, the degree of admixture is different when comparing these 2 SCD populations from eastern and western regions of the US. Overall, SCD patients in North America have a high degree of genetic admixture compared to patients with SCD from other regions of the world. The degree of admixture adjustment could vary for SCD subjects from different geographic regions of the US. Finally, these SCD populations may be ideal for mapping by MALD to identify genetic modifiers of this monogenic disease.

2932F
Erythropoiesis failure and ribosomal dysfunction in zebrafish model of Diamond-Blackfan anemia. N. Kenmochi1, T. Uechi1, T. Uechi1, Y. Nakajima1, G. Yada1, T. Sawada1, M. Ikeda2. 1) Frontier Sci Res Ctr, Univ Miyazaki, Miyazaki, Japan; 2) Veterinary Pharmacology Dept, Univ Miyazaki, Miyazaki, Japan.
Ribosomes, the molecular factories that carry out protein synthesis, are essential for every living cell and ribosomal proteins (RPs) play important roles in the formation of a functional ribosome. Defects in ribosome biogenesis have been linked to many human diseases called ribosomopathies, a rare collection of genetic disorders that are associated with increased cancer susceptibility. Diamond-Blackfan anemia (DBA) represents the first and the most extensively studied human disease caused by defects in ribosomal proteins. RPS19 is most commonly mutated in DBA, although some patients show mutations in several other RPs. To investigate the molecular pathogenesis of DBA, we have developed a zebrafish model of DBA by knocking down the <italic>italic rps19</italic> gene in the embryos using a morpholino oligonucleotide (MO). The knockdown embryos displayed a drastic reduction of red blood cells, whereas differentiation of other myeloid cells and endothelial cells seemed to be normal. The anemia phenotype was almost completely rescued by injection of synthesized <italic>rps19</italic> mRNA, but not by mutated mRNAs with patient-type mutations. The DBA model also showed developmental abnormalities in the head and tail regions due to increased cellular apoptosis. Simultaneous inhibition of <italic>p53</italic> rescued the morphological abnormalities but did not alleviate the erythroid aplasia, suggesting that a <italic>p53</italic>-independent but <italic>rps19</italic>-dependent pathway could be responsible for defective erythropoiesis in DBA. To evaluate the impact of <italic>rps19</italic> deficiency on mRNA translation, we carried out polysome analysis. The polysome patterns were similar between <italic>rps19</italic> morphants and control embryos but the amount of heavier fractions was less in the morphants. Finally, we carried out an RNA-Seq analysis of polysomal mRNAs that were purified from the embryos. We found that translational efficiencies of about 30 mRNAs were significantly changed in <italic>rps19</italic> morphants. These data suggest that selected mRNA translation may play an important role for the pathogenesis of DBA.
A family based exome sequence analysis identifies a rare AID deficiency causing mutation enriched in Finland. L. Trotta1, H. Alnusa1, M. Lepisto1, P. Ellonen1, S. Hannula1, A. Polode1,2, K. Porkka1, M. Seppanen1, J. Saarela1. 1. Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 3) Hematology Research Unit Helsinki, Department of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 4) Immunodeficiency Unit, Division of Infectious Diseases, Helsinki University Central Hospital, Hospital District of Helsinki and Uusimaa, Aurora Hospital, Helsinki, Finland.

Primary immune deficiencies (PIDs) are a large and heterogeneous group of congenital disorders of the immune function due to genetic defects impairing essential branches of the immune system. PIDs predispose affected individuals to increased susceptibility to recurrent and persistent infections and autoimmunity. PIDs are mainly inherited as rare monogenic disorders with more than 200 different genetic etiologies (5670 known mutations in genes with immune function) occurring in ~1:500-10,000 live births, with differences among countries and a much higher prevalence of some disorders in cases with high consanguinity rates or among genetically isolated populations. The molecular mechanisms behind many forms are not yet known and the significant clinical and immunological heterogeneity often delay the diagnosis and make treatment of PIDs challenging. Several cases of PIDs have been described in Finland, and because its unique genetic architecture the Finnish population has been defined as a model population for human genetic studies, especially for monogenic disorders. A founder effect, drifts in subsequent subisobates and rapid regional expansions shaped the gene pool of the population leading to the peculiar enrichment of rare variant responsible for the occurrence of the ‘Finnish disease heritage’ (i.e. more than 40 rare genetic disorders enriched in Finland). An exome sequencing analysis of a Finnish PID family with 5 affected siblings, characterized by lymphatic hyperplasia, autoimmunity and recurrent respiratory tract infections, as well as missing IgA, IgG, IgE levels and switched memory B cells, led us to identify the p.Met139Thr mutation in the AICDA gene. The gene, previously reported to cause AID deficiency, encodes for a single-stranded DNA-specific cytidine deaminase required for several crucial steps of B-cell terminal differentiation necessary for efficient antibody responses. The identified mutation is a rare variant, not reported in the 1000 Genomes database. We further monitored the frequency of the mutation in ~4000 Finnish individuals with sequence level data (kindly provided by the SuSu Consortium) observing it as very rare (0.152%), however 12 fold higher in Finland than in the general European population (NHLBI Exome Sequencing Project data). This suggests an enrichment of the variant in the Finnish population and further studies may provide evidence to merit the disease to be included in the list of ‘Finnish disease heritage’.

Exome sequencing identifies NFKB2 mutations as a cause of autosomal dominant early-onset common variable immunodeficiency. K. Chan1, E. Coonrod2, A. Kumanovic3,5, Z. Franks2, J. Durtche2, R. Marggraf2, W. Wu3, N. Augustine3, P. Ridge2, H. Hill2,5, 4, L. Jorde6, A. Weyrich7, G. Zimmerman1, J. Bohnsack1, K. Voelkerding2. 1) Dept. of Pediatrics, Allergy & Immunology, University of Utah School of Medicine, Salt Lake City, UT; 2) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 3) Dept. of Medicine, University of Utah School of Medicine, Salt Lake City, UT; 4) Dept. of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT; 5) Dept. of Pathology, University of Utah School of Medicine, Salt Lake City, UT; 6) Dept. of Human Genetics, University of Utah, Salt Lake City, UT; 7) The Molecular Medicine Program, University of Utah, Salt Lake City, UT.

Common variable immunodeficiency (CVID) is a heterogeneous disorder involving low antibody levels, poor response to antigens, and immune dysregulation. CVID is one of the most common primary immune deficiencies, occurring in approximately 1:10,000 to 1:25,000 people. The genetic defects responsible for CVID have been identified in less than 15% of cases. To investigate the molecular cause of CVID, we studied a family affected by an autosomal dominant form of early-onset CVID. Exome sequencing analysis of four members of this family identified a heterozygous frameshift deletion (c.2563delA; p.Lys855SerfsX7) in the gene that encodes nuclear factor kappa-B2 (NFKB2; p100/p52, also known as p100/p49). This mutation is predicted to both truncate the protein and prevent nuclear translocation due to impairment of phosphorylation and proteosomal processing of the C-terminal into the active protein. Using long-range PCR and next-generation sequencing, we then screened for NFKB2 mutations in a cohort of 33 additional CVID patients and identified a second heterozygous nonsense mutation at the same C-terminal locus (c.2557C>T; p.Arg853X) in an unrelated CVID patient. All four patients from the two families with NFKB2 mutations presented with an unusual combination of childhood-onset CVID and adrenal insufficiency. Variant Annotation, Analysis, and Search Tool (VAAST) analysis of the exome sequence data from the index family and an unrelated CVID patient ranked NFKB2 highest and reached genome-wide significance (p< 1.56x10-11). Western blotting and immunocytochemical analysis of NF-kappa-B2 in affected subjects demonstrated the presence of mutant NF-kB2 protein and decreased nuclear translocation. NF-kB2 is the principal protein involved in the noncanonical NF-kB pathway, is evolutionarily conserved, and functions in peripheral lymphoid organ development. B cells are the primary antibody producing site. The identification of NFKB2 mutations will further elucidate the pathogenic mechanisms of CVID and can provide potential targets for treatment. These findings describe the first primary immunodeficiency syndrome caused by a germline mutation in the noncanonical NF-kB signaling pathway.
Lobar holoprosencephaly (HPE) associated with additional clinical anomalies in two daughters of a consanguineous couple. Comparison of SNP-array analysis results as an attempt to search for a potential causative candidate gene. P.M. Kroisel1, B. Piecko2, M. Brunner-Krainz1, M.R. Speicher3, C. Windpassinger1. 1) Human Genetics, Medical University of Graz, Graz, Styria, Austria; 2) Department of Pediatrics, University of Zurich, Zurich, Switzerland; 3) Department of Pediatrics, Medical University of Graz, Graz, Austria.

A consanguineous family (both parents are first degree cousins) has five children, three boys and two girls. A lobar holoprosencephaly (HPE) was already diagnosed prenatally in both female pregnancies. Male offspring are healthy, however one of them shows a mild form of UDP-galactose-4-epimerase deficiency. In the first affected girl by prenatal sonographic investigation HPE was suspected in the 23. gestational week and was confirmed by MRI in the 26. gestational week. Cytogenetic analysis and array CGH performed after birth at term showed normal results but distinct facial features like hypertelorism, a depressed nasal bridge, low set simplified ears and a general increased skin hair density and length were recognized. Beside brain anomalies like the mentioned lobar HPE, a ventriculomegaly and a partial pachygyria was found. Psychomotor development was strongly reduced, and seizures, which became more frequent and severe, finally caused a fatal outcome at an age of 3 years. Her sister, who was born two years later shows a nearly identical clinical phenotype with lobar HPE, very similar facial and skin hair anomalies but also a hypoplastic vermis. Seizures are however clearly less frequent and antiepileptic treatment is more successful. However progress in her neurological development is of course related to her brain anomalies very severely delayed. Cytogenetic and array CGH analysis using a 60k Agilent oligonucleotide array also did not reveal a genomic unbalance or aberration. Based on these findings a monogenic form of HPE appears to be likely in our two patients. At least 14 genes have been implicated in HPE, with 4 major genes including SS and 10 minor genes. Yet still had sufficient DNA of the deceased patient and her severely affected sister, who up to now seems to have a slightly better prognosis. Therefore we performed a 250k Affymetrix SNP-array analysis of both DNA-samples in an attempt by homozygosity mapping and comparison of the results in all available affected and non-affected regions. By this approach we were able to reduce the number and size of homozygous chromosomal segments shared by both sisters to two chromosomal areas of less than about 2-3 Mb at 10p11 and 11q24. Only to the latter location the candidate gene EEDON has been mapped. If analysis of this gene would fail to show a homozygous mutation, NGS analysis could be considered to identify a possible new causative gene for HPE.
INTRODUCTION: Prune belly syndrome (PBS) is a rare complex of abnormal abdominal wall musculature, urinary tract anomalies, and bilateral cryptorchidism, and is associated with significant mortality. The cause of PBS is unknown, however a male predominance (95%) and identified families with multiple affected individuals suggests PBS may be an X-linked or sex-limited genetic disease. Disease-causing DNA changes can vary from chromosomal deletions/amplifications to being caused by known genetic markers that can be identified by DNA sequencing. More recently discovered, copy-number variations (CNVs) are small deletions or duplications that affect the number of copies of a gene. CNVs are missed by DNA sequencing and karyotyping but have been shown to account for previous unexplained genetic diseases by identifying candidate genes or regions involved in various disorders. The purpose of this study was to assess if novel CNVs are present in PBS patients. MATERIALS & METHODS: From 2008-current, blood samples from PBS patients were prospectively tested by whole genome comparative genomic hybridization (CGH) using the array version current (V8.1 or V8.3) at the time of ascertainment (Baylor Medical Genetics Laboratory, Houston TX). When possible, in cases wherein novel genomic CNVs were detected, parental DNA was also tested to identify whether CNVs are de novo or inherited. RESULTS: CGH testing on 12 PBS cases identified 5 (42%) patients with novel CNVs. Two patients have duplication CNVs: one patient has a maternally inherited 0.360Mb duplication on 2q11.2 and one patient has a 0.324Mb duplication on Xq23 of unknown genetic origin. Two patients have deletion CNVs: one patient has a maternally inherited 0.271Mb deletion on 7q31.1 and one patient has a maternally inherited 0.113Mb deletion on Xq22.1. Additionally, one adopted patient has two 0.066-0.151Mb deletions on 6q23.2 and 7q35.2 of unknown genetic origin. CONCLUSIONS: Prune belly syndrome (PBS) is a rare but morbidity disease and novel CNVs are common in these males (42%). The finding that 2 of the 5 identified CNVs are located on chromosome X supports the hypothesis that PBS is an X-linked genetic disease. Further investigation of these genomic rearrangements may lead to the identification of genetic causes of PBS, thereby aiding prenatal diagnosis and genetic counseling.

2940T
A Mutation at the H2B Histone Family, member W [H2BFWT] gene causes a novel X-linked mental retardation with abnormal head shape syndrome. V. Chini1, R. Ali4, N. Khattab1, T. Bin Omran2, Y. Al-Sarraj1, M. Kambouris1,2, H. El-Shanti1,2,4, 1) Molecular Genetics, Shafallah Medical Genetics Center, Doha, Qatar; 2) Pediatric Department, Hamad Medical Corporation, Doha, Qatar; 2) Yale University School of Medicine, Genetics, New Haven CT, USA; 4) University of Iowa, Pediatrics, Iowa City, IA, USA.

A non-consanguineous Arabic family affected by a putative novel seemingly X-linked disease characterized by mental retardation and abnormalities in head shape were studied by gene mapping, candidate gene mutation screening and whole X chromosome Exome sequencing of a single affected member to identify the responsible gene defect. Clinical presentation includes mental retardation, hyperactivity, hypotonia, turricephaly (in one affected), dolichocephaly (in the other affected), narrow face, downward slanted palpebral fissures, large ears, open mouth appearance, long and slender fingers. Neurologic and metabolic evaluations including urine organic acid, lysosomal enzyme analysis, gaudinicus, CDG and Fragile-X syndrome were performed. SNP genotyping with the HumanOmniExpress bead chip [Illumina, USA; analyzed with the GeneMapper software mapped the responsible gene to four possible X-chromosome intervals [p22.23-22.2, p22.2-21.1, p31.1-22.3, q26.3-28] as the family structure did not allow identification of a single interval with a significant LOD score. Based on the clinical presentation one candidate gene (FGF16) was screened but no pathogenic mutations were identified. Whole Exome target enrichment Next Generation Sequencing for the X chromosome was performed on the ABI SOLID4 for a single affected individual. Three variants were identified within the linkage intervals: AMMECR1 [c.C208T:p.L70F] was excluded as it did not co-segregate with the disease phenotype. A second variant (c.C11T:p.P4L) in the Rab40 GTP-binding protein gene (RAB40AL). Pathogenic mutations within this gene have been associated with Martin-Probst Syndrome (MRXSMP) which presents with a different phenotype characterized by hearing loss. The last variant (c.G227A/p.C76Y) at the H2BFWT histone family gene has damaging effects according to PolyPhen and SIFT protein-modeling software, co-segregates to the disease phenotype and is absent in 752 ethnically matched control chromosomes. No known diseases have been associated with mutations in H2BFWT apart from male infertility. Tissue specific H2BFWT expression studies show expression in human fetal brain, making this mutation the likely developmental defect. At present, mutation analyses in additional healthy normal male relatives are underway for the possible exclusion of H2BFWT as the offending gene.

2941F
The Solute Carrier SLC26A9 Accounts for Variability in Biomarkers of Cystic Fibrosis-Related Prenatal Exocrine Pancreatic Damage. M.R. Miller1, D. Soave2,6, W. Li3,6, J. Chang2, J. Gong1, H. Levy2,4, L. Sun2, J.M. Rommens1,6, F. Accurso1,6, P. Durie1,6, M.K. Song4, L.J. Strug1,2, 1) Research Institute, The Hospital for Sick Children, Toronto, ON, Canada; 2) Biostatistics, University of Toronto, Toronto, ON, Canada; 3) Pediatrics, Medical College of Wisconsin, Milwaukee, WI; 4) Clinical Research Institute, Children's Hospital of Wisconsin, Milwaukee, WI; 5) Molecular Genetics, University of Toronto, Toronto, ON, Canada; 6) Pediatrics, University of Colorado Denver School of Medicine, Aurora, CO; 7) Pediatrics, Children's Hospital Colorado, Aurora, CO; 8) Pediatrics, University of Toronto, Toronto, ON, Canada; 9) Epidemiology, Colorado School of Public Health, Aurora, CO.

In Cystic Fibrosis (CF), exocrine pancreatic damage begins in utero. Serum levels of the pancreatic enzyme trypsinogen (IRT) are elevated at birth in CF individuals and form the basis of newborn screening. Postnatally, IRT levels decline rapidly, and reflect pancreatic reserve. IRT levels at birth act as a biomarker of prenatal exocrine pancreatic damage, and are heritable and variable, even in individuals with the same CFTR genotype, indicating a role for modifier genes. On average, IRT is lower in CF individuals with the intestinal obstruction Meconium ileus (MI). In ~6300 CF individuals from the International CF Gene Modifier Consortium (the Consortium), eight SNPs in three solute carriers, SLC26A9, SLC6A14 and SLC9A3, were associated with MI, as was a prioritized gene set corresponding to apical plasma membrane constituents that reside alongside CFTR. We hypothesized that these genetic contributors to MI would be associated with exocrine pancreatic damage, as measured by newborn screened (NBS) levels of IRT. Genome-wide association studies (GWAS) of IRT were conducted in adults. Thereafter, we sought to determine whether any of the eight SNPs in SLCS26A9, SLC6A14 and SLC9A3 were individually associated with IRT, as is done in CFTR, but all four tested SNPs in SLCS26A9 were significantly associated with IRT after multiple test correction (SNP rs7512462, adjusted p = 3.82 x 10-4). Each MI risk allele is associated with ~53 ng/ml reduction in IRT (95% CI, -81, -28); this SNP explains 11% of the variation in IRT. While not significant at the genome-wide level, the GWAS-HD identified rs7512462 in SLC26A9 as the top-ranked SNP. The GWAS-HD suggested SLC26A9 plays the greatest role in exocrine pancreatic damage among all tested constituents of the apical membrane, demonstrating that biologically prioritizing genome-wide data can be a powerful tool in association studies. Together with the previous associations with MI, CF-related diabetes, and reported interaction with CFTR, we propose SLC26A9 is a therapeutic target to ameliorate CF disease severity.
Identification of a novel locus on chromosome 1 for autosomal dominant retinitis pigmentosa. K.K. Selmer1,*, R. Riese2, M.D. Vigeland1, K. Brandal3, D.E. Undlien1. 1) Department of Medical Genetics, Oslo University Hospital and University of Oslo; 2) Department of Ophthalmology, Innlend Hospital, Elverum, Norway.

Purpose: The aim of this study was to identify the genetic cause of autosomal dominant retinitis pigmentosa in a large Norwegian family.

Methods: We ascertained 20 members of a Norwegian family with autosomal dominant retinitis pigmentosa in three generations. Maximum possible LOD score was estimated in easyLINKAGE. Peripheral venous blood was drawn and clinical ophthalmological examination performed. We genotyped all 20 individuals using Affymetrix GeneChip® Human Mapping 10K 2.0 array and performed linkage analysis in the software MERLIN. Candidate genes were sequenced using traditional Sanger sequencing.

Results: Assuming full penetrance and no phenocopies, the maximum possible LOD score of this family was estimated to be 3.6. Multipoint linkage analysis revealed a single region achieving this maximum score. The linkage peak spans 5 Mb on the short arm of chromosome 1, and contains more than 100 protein coding genes, of which none are known to cause retinitis pigmentosa. Two candidate genes, the SLC6A1 and SLC6A4 were sequenced, but no mutation was identified.

Conclusion: We have identified a new locus for autosomal dominant retinitis pigmentosa on chromosome 1. The linked region is 5 Mb long and contains more than 100 genes. The sequencing of two candidate genes did not lead to the identification of the disease gene.
Novel frameshift mutation of the ADAR1 gene in a Chinese family with dyschromatosis symmetrica hereditaria. W. Cao 1, J. Zhang 1, X. Chen 1, Z. Zhang 2, H. Weng 2, L. Guo 1, Y. Liu 1, Z. Zhao 1, Q. Xing 1, M. Shao 1, 1) Clinical Research Center, Zhengzhou People’s Hospital, Zhengzhou, China; 2) Children’s Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai, China; 3) Ninth People’s Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; 4) Department of Urology, People’s Hospital of Henan Province.

Dyschromatosis symmetrica hereditaria (DSH; MIM127400) is a human pigmentary genodermatosis which has been reported predominantly in East Asians. This condition is typically present with a mixture of hypopigmented and hypopigmented macules on the dorsal aspects of the extremities and freckle-like macules on the face. The DSH locus has been mapped to chromosome 1q21.3, and pathogenic mutations were identified in the ADAR1 gene. ADAR1 catalyzes the conversion of adenosine to inosine RNA editing on double-stranded RNA substrates. Human ADAR1 protein includes two translation products due to alternative initiation of transcription: a full-length 150-kDa ADAR1p150 and a shorter 110-kDa ADAR1p110, which is initiated from a downstream AUG at codon 296. In this study, a family with autosomal dominant DSH from Jiangsu province of China was investigated. The pedigree comprised of 13 individuals spanning three generations. The proband was a 9-year-old boy who presented typical hyperpigmented and hypopigmented macules that varied in shape and size on the dorsal aspects of his hands and feet since the age of 5. Direct sequencing of the patient’s ADAR1 gene revealed the presence of a novel heterozygous deletion mutation c.271-272delAG (p.R91fsX123) in exon 2. This mutation leads to a frameshift which results in a premature translation peptide consisting of 151 amino acids without translation donut and creates a premature translation stop codon at codon 296, thus cause early stopped ADAR1p150 synthesis in the patients carrying these mutations, but there should be no effect on the synthesis of the ADAR1p110 proteins. Moreover, using luciferase reporter assay, we confirmed that the identified novel mutation has no effect on the synthesis of the p110 protein, although it located in the 5’-UTR of p110. Our data confirmed the dosage of functional p150 is the determinant of DSH.


Objective: The aim of this case-control study was to determine whether Single Nucleotide Polymorphisms (SNPs) located within or near the Activin-B and Activin-B genes were associated with non-3rd molar Naturally Missed Tooth (NMT) and oropharyngeal cleft palate at 5.5 and 14.5 ± 4.6 years, respectively. Clinical assessment of the dentition, dental histories and radiographic verification were utilized to accurately identify cases of tooth agenesis versus teeth that were missing due to other causes (extractions due to caries, trauma, or orthodontic treatment). Subjects Research. Genomic DNA was isolated from the saliva of 97 orthodontic patients and controls. The mean age ± standard deviation for the subjects with NMT and controls were 16.0 ± 4.6 years. Single Nucleotide Polymorphisms (SNPs) located within or near the Activin-B genes were assayed using Taqman®-based SNP genotyping. Hardy-Weinberg Equilibrium (HWE) was assessed based on SNPs with Hardy-Weinberg equilibrium. A novel SNP rs7576183 A-allele predominately presented with incisor hypodontia of the maxillary arch. Conclusion: Each rs7576183 A-allele provides a multiplicative reduction of 0.41 (0.17, 0.90) in the odds of hypodontia. Conversely, each copy of the G-allele at rs7576183 confers a 2.44 (1.11, 5.88) increased odds of hypodontia.

New candidate genes in holoprosencephaly: results from homozygosity mapping in six inbred families. S. Odent 1,2, M. de Tayrac 1,2, M. Bab 2,3, A. C. Dubourg 2,3, C. Mousselin 1, R. Bouvet 1, S. Guézou 2, M. Saha boulo 1, L. Ratié 1, V. Dupé 1, J. Mosser 2,3,6, V. David 2,3, L. Genetique Clinique, CHU de Rennes Hosp Sud, Rennes CDX 2, France; 2) UMR 6290 CNRS, IGG, University Rennes1, Rennes, France; 3) Service de génétique moléculaire et génétique, CHU de Rennes, Rennes, France; 4) Inserm UMR 946, Genetic variability and human diseases, Paris, France; 5) Institut Universitaire d’Hématologie, Université Paris Diderot, Paris, France; 6) Plateforme Biogenouest, Rennes, France; 7) Fondation Jean Dausset CEPH, Paris, France.

Holoprosencephaly (HPE) is a congenital malformation of the human brain due to an imperfect division of the forebrain during early development. Multiple genetic defects have been identified as involved in this process. It is now currently admitted that HPE is a multifhit pathology caused by at least two or more dysfunctional events involving at least 4 major genes (SHH, ZIC2, SIX3 and TGIF) and 10 minor genes belonging to different signaling pathways. However, the mutations and deletions in these genes represent only 30% of HPE cases. Recessive inheritance of HPE can also be suspected in consanguineous families with intrafamilial recurrence. Homozygosity mapping was undertaken in six families with history of consanguineous marriage to search for regions harboring mutations that are identical by descent. Parents and affected children were genotyped on HumanCytoSNP-12 arrays (Illumina). We first determined the population ancestry of each family with Origiminer [de Tayrac, ASHG 2012] to estimate their specific SNP allele frequencies. Inbreeding coefficients of affected children were estimated from their genomic data by the FEstim method [Leutenegger et al. 2012]. Using inbreeding and homozygosity mapping, we performed homozygosity mapping without relying on the genealogical information. In parallel we detected the runs of homozygosity across the genome for each individual (PLINK). Three regions of interest were detected by both methods on chromosomes 6 (D6S1173), 7 (D7S1601) and 10. We applied Endeavor software to prioritize genes in these regions. One of them was previously identified by CGH array on chromosome 6 (D6S1173) and other good candidate genes will be presented.
suggest that this interaction might explain the mutant's selective advantage
STAT3. Given STAT3's function in mouse spermatogonial stem cells, we
and autosomal recessive inheritance has been described only rarely.
In this study, we analyzed whole-exomes of 18 probands with heterotaxy-
This includes 4 probands belonging to families with
obvious autosomal recessive inheritance. Exome analyses identified approx-
imately ~8000 coding variants in each individual which were subsequently
filtered using 1000 Genome, dbSNP135 and NIH exome sequencing project
(ESP) databases. The remaining variants were further analyzed using
regions of homozygosity (ROH) revealed by SNP genotyping and/or inher-
ance-based exome analysis. This identified suspected disease-causing var-
iants in novel candidate genes in 7 probands as well as in previously known heterotaxy or related phenotype genes in 4 probands. The majority of var-
ants were homozygous and located in ROH (~2Mb-44Mb), supporting
the autosomal recessive inheritance in 6 familial cases. Sanger sequencing of probands and available family members confirmed these homozygous
mutations and validated their recessive segregation with phenotype. Novel
selected candidates (C2orf59, DNNAN, DNAAB1, MIAS, INST7, JMJD1C, SORBS2) were further investigated by expression analyses in mouse, including RT-PCR and whole mount in-situ hybridization, with results sug-
gest a possible role in LR patterning and early heart development. Knock-
down of c2orf59 via antisense morpholin in Xenopus laevis resulted in
LR patterning defects and abnormal pilx2 expression at stage 30, suggesting
c2orf59 is a novel gene involved in laterality. This study proves the impor-
tance of exome analysis complemented with ROH segregation revealed by
SNP arrays for identification of unknown causes of human laterality defects
and these causes will provide insights to understand new molecular mecha-
nisms involved in embryonic LR patterning.

Age-Dependent Germline Mosaicism of the Most Common Noonan Syndrome Mutation Shows the Signature of Germline Selection. P. Calabrese1, S. Yoon1, S. Choi1, J. Eberwine1, B. Selb2, N. Arheim1, 1 Mol Comp Biol, Univ Southern California, Los Angeles, CA; 2) Icahn School of Medicine at Mount Sinai, New York, NY.

Noonan syndrome (NS) is among the most common Mendelian genetic diseases (~1/2,000 live births). Most cases (50%-84%) are sporadic, and new mutations are virtually always paternally derived. More than 47 different sites of NS de novo missense mutations are known in the PTPN11 gene that codes for the protein tyrosine phosphatase SHP-2. Surprisingly, many of these mutations are recurrent with nucleotide substitution rates substantially greater than the genome average; the most common mutation, c.922A>G, is at least 2,400 times greater. We examined the spatial distribution of the c.922A>G mutation in testes from 15 unaffected men and found that the mutations were not uniformly distributed across each testis. This would be expected for a mutation hot spot but were highly clustered and showed an age-dependent germline mosaicism. Computational modeling that used different stem cell division schemes confirmed that the data were inconsis-

tent with hypermutation, but consistent with germline selection. Mutations driven by germline selection in different gene families can help highlight common causal signaling pathways.

Mutations in Sporadic and Familial Congenital Diaphragmatic Hernia Patients. M. Longoni 1,2, M.K. Russell-2, F.A. High1,2, K. Lage1,2,3, J. Wells1, C.J. Bull2, K.G. Ackerman3, C. Lee4, B.R. Pober1,2, P.K. Donahoe1,2, 1) Massachusetts General Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Children's Hospital Boston, Boston, MA; 4) Broad Institute, Cambridge, MA; 5) The Jackson Laboratory, Bar Harbor, ME; 6) School of Medicine and Dentistry, University of Rochester, Rochester, Rochester, NY; 7) Brigham and Women's Hospital, Boston, MA.

Congenital Diaphragmatic Hernia (CDH) is a common birth malformation. CDH is genetically heterogeneous, and evidence points to a multifactorial, possibly polygenic, etiology in many cases. The most prevalent phenotype is isolated left-sided posterolateral hernia, associated with life threatening pulmonary hypoplasia. An intragenic deletion in Zinc finger protein, multitype 2 (ZFPM2, also known as FOG2), a gene previously implicated in CDH, was identified in a multigenerational family with isolated hernia inherited in an autosomal dominant pattern with incomplete penetrance. Point mutations in ZFPM2 or in its transcriptional co-repressor C-terminal binding protein (CTBP2) were identified in 7 unrelated patients from a cohort of 93 CDH patients. Exome sequencing from an unaffected parent, consistent with the hypothesis that multiple environmental factors or genetic modifiers are necessary for ZFPM2 haplosufficiency to display a clinically recognizable diaphragmatic defect. In an effort to identify pathogenic mechanisms of ZFPM2 mutations resulting in CDH, microarray expression data from the primordial diaphragm of Zfpm2 null mice at E11.5 were generated from laser captured tissue. The dataset, containing 428 differentially expressed genes (mutant vs. wildtype), was enriched for Mouse Gene Informatics (MGI) Mammalian Phenotype ontology categories such as skeletal, muscle, and respiratory phenotypes. This expression dataset, together with an analogous one from wild type mice collected at three time points during diaphragm development published by Russell and colleagues (2011), was used as a screening tool to identify possible second hits that explain the observed variable penetrance. In most patients with ZFPM2 deletions or point mutations, we identified additional novel or rare sequence variants, defined as less than 0.1% in control cohorts, mapped to genes deregulated in Zfpm2 knock-out animals or in genes whose expression levels are highly correlated with Zfpm2 during embryogenesis. Though further studies are needed to determine that our prioritized second hits are bona fide Zfpm2 genetic interactors, our results show that developmentally oriented expression studies are a useful tool for the interpretation of exome sequencing data when studying birth defects.
Whole exome sequencing links TMCO1 with Cerebro-Facio-Thoracic Syndrome. Y. Bayram1, E. Karaca1, D. Pehlivan1, C.R. Beck1, C. Gonzaga-Jauregui1, T. Gambin1, S.N. Ijhangari2, H. Aydin3, W. Witzniewski4, A.H. Celtib, M.M. Atlid, D. Muzny5, R.A. Gibbs5, J.R. Lupski5,6,7,8,9. 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston TX, USA; 3) Department of Medical Genetics, Zeynep Kamil Women’s and Children’s Hospital, Istanbul, Turkey; 4) Department of Medical Genetics, Karadeniz Technical University, Faculty of Medicine, Trabzon, Turkey; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 6) Texas Children’s Hospital, Houston, TX, USA.

Next generation massively parallel sequencing, both Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES), is a disruptive technology that has changed the course of clinical/human genetics. As a more efficient and cost-effective method, WES is now widely used as a diagnostic tool for identifying the underlying gene, mutation(s), and genetics of challenging genetic syndromes. Here we report a case with a clinical diagnosis of Cerebro-Facio-Thoracic Syndrome (CFTS) (MIM#213980) in whom we identified by WES homozygous splice site mutations in the transmembrane and coiled-coil domains of TMCO1 gene. Mutations in TMCO1 gene were shown to be responsible for craniofacial dysmorphism, skeletal anomalies, and mental retardation syndrome (MIM#614132), which is characterized by facial dysmorphism, multiple malformations of the vertebrae and ribs, and mental retardation. Our retrospective review revealed that clinical manifestations of both syndromes are remarkably similar. We propose that, mutations of TMCO1 are not only responsible for craniofacial dysmorphism, skeletal anomalies, and mental retardation syndrome, but also for CFTS. In addition, one should keep in mind that these two syndromes display different spectrum of the same clinical entity, molecularly diagnosed as TMCO1 defect syndrome.

Stormorken syndrome: mutation in STIM1 as a cause of a remarkable coincidence? G. Morin1, G. Jedraszak1, A. Rabbid singh2, B. Roméo2, E. Bourges-Petit3, H. Sevestre4, D. Brémont-Gignac5, H. Ouadid-Ahmdouch5, M. Mathieu1, J. Rochette1. 1) Genetic department, EA 4666, Amiens University Hospital, Amiens, France; 2) Pediatric cardiology unit, Amiens University Hospital, Amiens, France; 3) Pathology service, Amiens University Hospital, Amiens, France; 4) Ophthalmo service, Amiens University Hospital, Amiens, France; 5) University of Picardie Jules Verne, UFR of Sciences, Laboratory of Cellular and Molecular Physiology EA 4667, SFR CAP-SANTE, Amiens, France.

Background Stormorken syndrome is a very rare autosomal dominant disease associated with congenital miosis, thrombocytopenia/thrombocytopathia, tubular aggregate myopathy, asplenia, small stature, ichthyosis, and moderate developmental disabilities. When documented the myopathy is characterized by the presence of tubular aggregates. Additionally, most patients present moderate hypocalcemia. Till now, only 4 families have been reported with a total of 6 patients. The moderate hypocalcemia observed in some patients suggested the deregulation of the calcium homeostasis. Furthermore, mutations in the calcium sensor region of STIM1 were reported in the autosomal dominant form of tubular aggregate myopathy. Methods We describe a family study expanding over 2 generations with 2 affected individuals displaying Stormorken syndrome. The disease affected father and son confirming its autosomal dominant inheritance. Both proband and son displaying Stormorken syndrome. The disease affected father and son, and not in the other members of the family. Restriction enzyme analysis confirmed the presence of the C>T substitution in position 910. The identified mutation lies in the coiled coil 1 domain of the protein (Cox2). This amino acid residue is highly conserved in mammals. Data obtained from in silico analysis show a major change in the secondary structure of the Cox2 domain when compared to the wild type.

FX10 in the development of the olfactory ensheathing cells, V. Pinault1,2,3, W. Watanabe1,2, S. Marcat1,2,3, Barat1,2,3, A. Chauoe1, M. Goosens1,2,3, J.P. Hardegen1, C. Dodé1,2, N. Bondurand2,3. 1) Genetics, INSERM U955 hospital Henri Mondor, Creteil, France; 2) Université Paris Est, Faculté de Médecine, Créteil, France; 3) Hôpital Henri Mondor, Laboratoire de Biochimie et Génétique, Créteil, France; 4) INSERM U1016, Institut Cochin, Département de génétique, Université Paris Descartes, Paris, France; 5) INSERM U587, Département de neuroscience, Institut Pasteur, Université Pierre et Marie Curie, Paris, France; 6) Laboratoire de biochimie et génétique moléculaire, APHP, Hôpital Cochin, Paris, France.

The SOX10 transcription factor plays a role in the maintenance of progenitor cell multipotency, lineage specification, cell differentiation, and is a major actor in the development of the neural crest. It has been implicated in Waardenburg syndrome (WS), a rare disorder characterized by the association of pigmentation abnormalities and deafness, sometimes associated with neurological disorders. We recently found that apart from WS, SOX10 mutations also cause Kallmann syndrome (KS) with deafness. KS is defined by the association of anosmia and hypogonadotropic hypogonadism due to incomplete migration of neuroendocrine GnRH (gonadotropin-releasing hormone)-cells along the olfactory, vomeronasal, and terminal nerves. We provide further study of SOX10 expression in the olfactory ensheathing cells (OECs) during mouse development and of the developmental defects observed in the Sox10 mutant mouse. This confirms several of the roles previously suspected for these peculiar cells, which are thought to be responsible for the olfactory neurons unique capacity to renew throughout life.
2959F
Exome sequencing in Mendelian cleft lip and palate families: Results of an International Orofacial clefting consortium. T. Roscioli1, M. Buckley2, C. Carels1, T. Cox1, E. Haa1, D. Hanna1, A. V. Hing1, K. Khandelwal1, E. Kirk1, A. Lidral1, J. Murray12, D.A. Nickerson3, F.Reiner1, J. Smith1, E. Thompson1, H. van Bokhoven11, H. Zhou12, 1) School of Women’s and Children’s Health, University of South Wales, Sydney, NSW, Australia; 2) Department of Haematology and Genetics, South Eastern Area Laboratory Services, Sydney, Australia; 3) Department of Orthodontics and Craniofacial Biology, Radboud University, Nijmegen, The Netherlands; 4) Department of Pediatrics, University of Washington, Seattle, USA; 5) South Australian Clinical Genetics Service, Adelaide, Australia; 6) The University of Washington Center for Mendelian Genomics, Seattle, USA; 7) Department of Pediatrics, Division of Craniofacial Medicine, University of Washington and Children's Hospital and Regional Medical Center, Seattle, USA; 8) Department of Medical Genetics, Sydney Children’s Hospital, Sydney, Australia; 9) College of Dentistry, University of Iowa, Iowa City, USA; 10) Department of Pediatrics, University of Iowa, Iowa City, USA; 11) Department of Human Genetics, Radboud University, Nijmegen, The Netherlands; 12) Department of Developmental Biology, Radboud University, Nijmegen, The Netherlands.

Cleft lip and palate is one of the most common birth defects, occurring in 1 of 700 live births. The majority of individuals occur in single affected member families. While GWAS studies have identified major loci, there remains an unexplained genetic contribution that we hypothesize is due to rare unidentified mutations. To this end, we have undertaken genomic studies in larger multi-affected families who could be enriched for Mendelian forms of orofacial clefting in order to identify causative high penetrance mutations in novel genes for which investigations are ongoing, including studies of the collagen family of genes and SOX9. We discuss in addition adjunct methodologies to reduce the number of variants requiring confirmation such as the programs homozgyosity mapper, dominant mapper and SNP arrays. In conclusion, based on this initial survey, we have identified multiple rare variants which may contribute to isolated cleft lip and palate.

2960W
Identification of a novel mutation Arg118Gly in the TWIST1 gene causing Saethre-Chotzen Syndrome. M. L. M. Castro1, C. H. P. H. Pavan-Lerpe1, R. R. Guarraseni1, C. M. Leprevost1, J. A. Josa-H-Kian1, L. F. Laureano2, W. A. R. Baratella3, J. Huber3, E. S. Ramos1, W. A. Silva Jr4, L. Martelli5, G. A. Molfetta1, 1) Serviço de Genética Médica - Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 2) Laboratório de Citogenética - Hospital Das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 3) Departamento de Genética - Faculdade de Medicina de Ribeirão Preto - Paulista, Ribeirão Preto - SP, Brasil; 4) Centro de Medicina Genética - HCFMRP - USP.

Saethre-Chotzen syndrome (SCS) or Acrocephalosyndactyly Type III is a craniosynostosis syndrome characterized by craniofacial and limb abnormalities resulting from premature fusion of variable exostoses. The disease exhibits autosomal dominant inheritance. Its prevalence varies between 1:25,000 and 1:50,000 live births. Since the first description of mutations in the TWIST1 gene causing SCS, about 163 different mutations, deletions and insertions have been detected in affected patients. The TWIST1 gene is located on chromosome 7p21.1 and encodes a transcription factor with a DNA-binding domain implicated in cell lineage determination and differentiation. We report on a novel missense mutation in the TWIST1 gene in a patient with SCS phenotype. The patient was referred to the Medical Genetics Division of the University Hospital when he was 7 months old and his dysmorphological examination showed: short stature, microbrachycephaly due to bilateral coronal synostosis, high forehead, low frontal hairline, flat face, midface hypoplasia, prominent eyes, unilateral eyelid ptosis, a broad and depressed nasal bridge, as well as hypoplastic, high arched palate, nystagmus, truncal and small head. Immunoblot demonstrates loss of this protein in the patient compared to controls. RT-PCR demonstrates that this splice mutation causes skipping of an exon in >13,000 American control DNAs and 200 Turkish control chromosomes.

In this study, we aim to elucidate the genetic cause of a rare recessive ataxia. Spinocerebellar ataxias with onset of symptoms occurring between early childhood and late adulthood. While the genetic causes of several ataxias have already been positively identified, many more, especially rare recessive forms, remain unknown. Heterogeneity, or the large number of genes that can cause recessive ataxias, is the major reason why many genes for recessive forms are still unknown. Individual families with independent mutations are too small to positively identify a chromosomal region by genetic linkage.

Objective: We are studying a consanguineous Turkish family with a non-progressive, congenital ataxia of unknown etiology. Symptoms include hypotonia, developmental delay, mental retardation, nystagmus, truncal and extremity ataxia and cerebellar hypoplasia that do not progress with age. In this study, we aim to elucidate the genetic cause of a rare recessive ataxia.

Methods: We utilized a combination of exome sequencing with homozygo- mosity mapping and expression analysis to identify candidate genes. Sanger sequencing was used to verify damaging variants. Molecular assays, such as RT-PCR and immunoblotting, were used to determine the functional consequences of gene variants.

Results: We have identified a variant in an obligatory splice sequence (the first base of an intron is changed from GT to AT). This mutation is absent in >13,000 American control DNAs and 200 Turkish control chromosomes. RT-PCR demonstrates that this splice mutation causes skipping of an exon that is present in all isoforms and expression of the gene is lower in patients than controls. Immunoblot demonstrates loss of this protein in the affected individuals.

Conclusion: Our results suggest that we have identified a novel gene causing recessive ataxia. Additionally, we are developing a zebrafish model, utilizing morpholino-mediated knockdown, to study this disorder. Our research may help identify causes of other ataxias and may lead to novel therapies to treat ataxias.

2961T
Splice site mutation leads to ataxia and retardation in a consanguineous family. R. M. Burns1, W. Peng1, J. Xu2, J. Z. Li2, M. Burmeister1, 1) Molecular and Behavioral Neuroscience Institute University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics University of Michigan, Ann Arbor, MI.

The cerebellar ataxias are a heterogeneous group of neurological disorders with onset of symptoms occurring between early childhood and late adulthood. While the genetic causes of several ataxias have already been identified, many more, especially rare recessive forms, remain unknown. Heterogeneity, or the large number of genes that can cause recessive ataxias, is the major reason why many genes for recessive forms are still unknown. Individual families with independent mutations are too small to positively identify a chromosomal region by genetic linkage.

Objective: We are studying a consanguineous Turkish family with a non-progressive, congenital ataxia of unknown etiology. Symptoms include hypotonia, developmental delay, mental retardation, nystagmus, truncal and extremity ataxia and cerebellar hypoplasia that do not progress with age. In this study, we aim to elucidate the genetic cause of a rare recessive ataxia.

Methods: We utilized a combination of exome sequencing with homozygo- mosity mapping and expression analysis to identify candidate genes. Sanger sequencing was used to verify damaging variants. Molecular assays, such as RT-PCR and immunoblotting, were used to determine the functional consequences of gene variants.

Results: We have identified a variant in an obligatory splice sequence (the first base of an intron is changed from GT to AT). This mutation is absent in >13,000 American control DNAs and 200 Turkish control chromosomes.

RT-PCR demonstrates that this splice mutation causes skipping of an exon that is present in all isoforms and expression of the gene is lower in patients than controls. Immunoblot demonstrates loss of this protein in the affected individuals.

Conclusion: Our results suggest that we have identified a novel gene causing recessive ataxia. Additionally, we are developing a zebrafish model, utilizing morpholino-mediated knockdown, to study this disorder. Our research may help identify causes of other ataxias and may lead to novel therapies to treat ataxias.
**2962F**

A novel gene for Spinocerebellar Ataxia (SCA) linked to chromosome 6 and involved in fatty acid metabolism. E. Di Gregorio1,2, B. Borroni1, L. G. Giorgio1, D. Lacerenza1, C. Mancini1, A. Calcia1, I. Mura1, D. Caviezzo2, N. Mitro3, M. Gaussen4, N. Io Buono1, A. Funaro1, G. Vaula5, I. Lagroua6, L. Orsi7, A. Durr4,6, C. Costanzi3, A. Padovani8, A. Brice6,8, L. Boccone9, E. Hoehn10, F. Tempia10, D. Caruso5, G. Stevanin10, A. Brusco10, 1) Medical Sciences, University of Torino, Turin, italy; 2) S.C.D.I Medical Genetics, Città della Salute e della Scienza, Torino, Italy; 3) University of Brescia, Department of Neurology, Brescia, Italy; 4) Laboratory of Human Genetics, Galliera Hospital, Genova, Italy; 5) University of Milano, Department of Pharmacological Sciences, Italy; 6) Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (INSELM/UPMC Univ. Paris 6, UMR, S975; CNRS 7225, EPHE), Pitié-Salpêtrière Hospital, Paris, France; 7) S.C.D.U. Neurology, Città della Salute e della Scienza, Torino, Italy; 8) APHP, Fédération de génétique, Pitié-Salpêtrière Hospital, Paris, France; 9) Ospedale Regionale Microcitemie, ASL 8, Cagliari, Italy; 10) University of Torino, Neuroscience Institute Cavalieri Ottolenghi (NICO), Torino, Italy.

Spinocerebellar ataxias (SCA) are a highly heterogeneous group of autosomal dominant neurodegenerative disorders phenotypically characterized by gait ataxia, incoordination of eye movements, speech, and hand movements, and usually associated with cerebellar atrophy. More than 30 SCA have been identified, whose genes may be classified into two main categories: repeat expansion disorders - among which the most common forms SCA 1, 2, and 6 - and genes with conventional mutations. Here we report the identification of a novel SCA gene. Genome-wide linkage analysis identified a 92 Mb region on chromosome 6 in an Italian family affected by a pure form of ataxia with disease onset in the 4th decade of life. Next generation sequencing of all coding genes in the smallest interval identified only one possible mutated gene, with a Gly to Val amino acid change. Screening of over 450 SCA independent patients identified this same mutation in two unrelated Italian families. Haplotyping proved that at least two of the three families shared a common ancestor. Two further missense variants, in three independent families, affected the same exon involved in the Italian families, suggesting this may be a mutational hot-spot. All changes hit conserved amino acids, and were not common polymorphisms. The gene encodes an ubiquitously expressed enzyme involved in fatty acid biosynthesis. In situ hybridization on mouse and human brain demonstrated that Purkinje cells have a peculiar high expression of the enzyme. In agreement with the function of this gene, we showed a reduced level of a subgroup of fatty acids in the serum of three patients with the Gly to Val amino acid change. Furthermore, we found an increased gene expression at the messenger RNA and protein levels in the patients' lymphoblasts vs. controls. We hypothesize that a positive feedback loop, activated by the functional impairment of the enzyme, may lead to an increase of protein expression that in turn may accumulate within cells leading to a toxic gain of function. In conclusion, we suggest that our mutated gene, highly expressed in Purkinje cells, is a good candidate for a new form of pure autosomal dominant cerebellar ataxia and join the group of CNS diseases involving fatty acids metabolism.

**2963W**

Targeted exome sequencing with copy number variant detection reveals HSD17B4-deficiency in a male with cerebellar ataxia and azosperma. D.S. Lieber1,2,3,4, S.G. Hershman1,2,3,4, N.G. Stefà1,2, S.E. Calvo1,2,3,5, K.B. Sims2,5, J.D. Schmahmann5, V.K. Mootha1,2,3,4, 1) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 3) Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA; 4) Broad Institute of Harvard and MIT, Cambridge, MA 02141, USA; 5) Department of Neurology, Massachusetts General Hospital & Harvard Medical School, Boston, MA 02114, USA; 6) Department of Medicine, Massachusetts General Hospital, Boston MA 02114, USA.

We describe an adult male with cerebellar ataxia, peripheral neuropathy, hearing loss, and azosperma. Commercial genetic testing of 18 ataxia and mitochondrial disease genes was negative, but biochemical findings in serum, urine, and muscle biopsy pointed to the possibility of a mitochondrial abnormality. Targeted exome sequencing followed by analysis of single nucleotide variants and small insertions/deletions failed to reveal a genetic basis of disease. We subsequently applied computational algorithms to infer copy number variants (CNVs) from exome data, revealing a heterozygous 12kb deletion of exons 10-13 of HSD17B4 that was compounded with a rare missense variant (p.A196V). Recessive mutations in HSD17B4 cause D-bifunctional protein deficiency, a severe, infantile-onset disorder of peroxisomal fatty acid oxidation. Recently, compound heterozygous mutations in HSD17B4 were reported in two sisters diagnosed with Perrault syndrome (MIM # 233400), who presented with ovarian dysgenesis, hearing loss, and azoosperma. Retrospective review of patient records revealed elevated ratios of pristanic:phytanic acid and arachidonic:docosahexaenoic acid, consistent with peroxisomal dysfunction. Our case expands the phenotypic spectrum of Perrault syndrome, representing the first male reported with ataxia, infertility, and HSD17B4 deficiency. Furthermore, our study highlights the importance of exome-based CNV detection in the diagnosis of rare disorders and points to potential crosstalk between mitochondria and peroxisomes in Perrault syndrome.
2964T
Mutations of COQ2 in Familial and Sporadic Multiple System Atrophy.
J. Mústila, T. Matsukawa, H. Ishiura, Y. Fukuda, Y. Ichikawa, H. Date, B. Ahsan, Y. Nakada, Y. Morose, Y. Takahashi, Y. Goto, Y. Momoto, K. Shiraihi, H. Takahashi, O. Onodera, M. Nishizawa, T. Kondo, S. Murayama, A. Dürre, A. Bride, A. Filla, T. Klockgether, U. Wüllner, G. Nicholson, S. Gilman, S. Tsuji, JAMSAC, JGSCAD, JPDDSG, ALNiMAC, NAMS-SG, 1Department of Neurology, Tohoku University of Kyoto, Tokyo, Japan; 2Department of Neurology, Tokyo University of Tokyo, Tokyo, Japan; 3Department of Molecular and Cellular Biosciences, The University of Tokyo, Tokyo, Japan; 4Department of Pathology, Brain Research Institute, Niigata University, Niigata, Japan; 5Department of Neurology, Brain Research Institute, Niigata University, Niigata, Japan; 6Department of Neurology, Wagakaya Medical University, Wagakaya, Japan; 7Department of Neuroepidemiology and the Brain Bank for Aging Research, Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan; 8Hôpital de la Salpêtrière, Département de Génétique et Cytophénétique, F-75013, Paris, France; 9Department of Neurological Sciences, University Federico II, Napoli, Italy; 10Department of Neurology, University of Bonn and German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; 11University of Sydney at the ANZAC Research Institute, Concord Hospital, Sydney, Australia; 12Department of Neurology, University of Michigan, Ann Arbor, MI.

Background Multiple system atrophy (MSA) is an intractable neurodegenerative disease characterized by autonomic failure with various combinations of parkinsonism, cerebellar ataxia, and pyramidal dysfunction. MSA is classified into two subtypes: MSA-C, characterized by predominant cerebellar ataxia, and MSA-P, characterized by predominant parkinsonism. MSA-C has been reported to be more prevalent (8.0 per 100,000) than MSA-P in the Japanese population, whereas MSA-P has been reported to be more prevalent in Europe and North America. Furthermore, multiplex MSA families have been identified. These findings strongly suggest involvement of genetic components underlying MSA. Methods In combination with linkage analysis, we performed whole-genome sequencing of an individual with autopsy-proven MSA of a multiplex family. Mutational analysis was further conducted on the patients in five other families, and, furthermore, on Japanese (363 MSA patients and 595 controls), European (56 MSA patients and 515 controls), and North American (172 MSA patients and 294 controls) series. Functional analysis of COQ2, which encodes para-hydroxybenzoate-polyprenyltransferase, an enzyme essential for coenzyme Q10 (CoQ10) biosynthesis, was conducted using a yeast complementation system and by enzyme activity measurement. CoQ10 levels in lymphoblastoid cells and brains were measured by high-performance liquid chromatography. Results We identified a homozygous mutation (M78V-V343A) and compound heterozygous mutations (R337X/V343A) in COQ2 in two multiplex families. Furthermore, we demonstrated that a common variant (V343A) and multiple rare variants of COQ2, both of which are functionally impaired, are associated with sporadic MSA. The V343A variant was exclusively observed in the Japanese population, while multiple rare variants are observed irrespective of ethnic background. The ratio of MSA-C to MSA-P was significantly higher in COQ2 variant carriers than in noncarriers. Conclusions Functionally impaired variants of COQ2 associated with an increased risk of MSA in multiplex families and sporadic cases, supporting a role of impaired COQ2 activities in the pathogenesis of MSA.

2966W
Autosomal-recessive spastic ataxias: systematic whole-exome sequencing of a large cohort reveals novel phenotypes and gene candidates. R. Schüle, M. Gonzalez, J. Reichbauer, A. Caballero Olaya, S. Wiethoff, M. Dobler, U. Gaisser, T. Rattay, K. Karle, S. Schols, S. Zuchner, M. Synofzik. 1) Department of Neurodegenerative Disease, Herit­tle Institute for Clinical Brain Research, Tuebingen, Germany; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida; 3) Brain Research Trust, UCL, Institute of Neu­rology, London, UK; 4) Department of Paediatric Neurology and Develop­mental Medicine, Children’s Hospital, University of Tuebingen, Tuebingen, Germany.

Autosomal recessive hereditary spastic ataxias are a clinically and geneti­cally poorly defined group of neurodegenerative disorders characterized by the combined occurrence of ataxic and spastic features. Patients are usually classified as either complicated Hereditary Spastic Paraplegia (HSP) or autosomal recessive spinocerebellar ataxia (SCAR), often with a large overlap which makes the respective classification arbitrary in many cases. Few genes are known to cause spastic ataxia as the predominant phenotype and they explain only a small fraction of known cases. Among the most common known causes of autosomal recessive spastic ataxias are mutations in the SACS gene, causing autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS), late onset Friedreich ataxia, caused by repeat expansions in the FXN gene and HSP Type 7 (SPG7). Using a comprehensive whole-exome sequencing approach, we aimed to systematically unravel the genetic basis of spastic ataxias, investigate atypical phenotypes and identify novel gene candidates. We have collected a cohort of 51 index patients with progressive early onset spastic ataxia and a family history of hereditary spastic and/or cerebellar ataxia. Mean age of disease onset was 10.8 years. Additional symptoms present in some of the patients included mental retardation, optic atrophy, peripheral neuropathy, epilepsy and dysgenesis of the corpus callosum on MRI. Whole-exome sequencing was performed in all index cases (enrichment: SureSelect Human All Exon 50Mb kit, Agilent; sequencing: Hiseq 2000, Illumina). After alignment to the hg19 version of the human genome (Wurrows-Wheeler algorithm) and variant calling (Genome Analysis Toolkit GATK software package) data was imported into GEM.app, a browser accessible exome database and collaboration tool. Pathogenic mutations in known genes were identified in about one third of cases These include mutations in known spastic ataxia genes like SACS, SPG7, GBA2, or FA2H, but also genes not yet associated with spastic ataxia like e.g. ADCS, SYNGAP1 or TAC. Clinico-genetic findings in solved families as well as exciting new candidate genes for autosomal recessive spastic ataxias will be presented. These findings not only provide a systematic account of the genetic basis and phenotypic spectrum of spastic ataxias, but also enlarge the genetic spectrum that needs to be considered in spastic ataxias.

1) Centre de génétique humaine, Université de Franche-Comté, Besançon, France; 2) Pathology Wellcome Trust Centre for Mitochondrial Research, New Medical School, University of Cambridge, Cambridge, UK; 3) Laboratory of Mitochondrial disorders University of Gent, Belgium; 4) CRICM (INSERM / UPMC 975, CNRS 7225) and neurogenetics laboratory, GHU Pitié-Salpêtrière, Paris, France; 5) Albrecht-Kossel-Institute for Neuroregeneration, University of Rostock, Rostock, Germany; 6) Mitochondrial DNA Replication Group National Institute of Environmental Health Sciences, Durham, NC; 7) Dept Neurology, Regional Hospital, Namur, Belgium; 8) Metabolic Unit, Centre de Human Genetics, University Hospital, Lige, Belgium; 9) Born-Bunge Foundation, University of Antwerp, Belgium; 10) Neurophysiology, private practice, Brussels, Belgium; 11) Dept Neurolgy, CHU UCL Mont-Godinne Dinant, Yvoir, Belgium and Institute of Neuroscience (IoNS), Université Catholique de Louvain (UCL), Brussels, Belgium.

Mutations in an increasing number of nuclear genes involved in the maintenance and replication of mitochondrial DNA (mtDNA) are being described, associated with an extensive spectrum of clinical phenotypes and secondary mtDNA defects that can take two forms: mtDNA depletion syndromes which are characterised by a quantitative loss of mtDNA copy number leading to isolated organ or multi-systemic pediatric mitochondrial disease and multiple mtDNA deletions which are associated with late-onset mitochondrial disease, a mosaic pattern of cytochrome c oxidase (COX)-deficiency and in which progressive external ophthalmoplegia (PEO) is the predominant clinical feature. We describe the clinical, pathological and molecular findings in a large Caucasian family segregating a late-onset (sixth to seventh decade) movement disorder comprising head and arms tremor, cerebellar ataxia, parkinsonism and sensory axonal peripheral neuropathy, complicated after a few years by generalized epileptic seizures and cognitive decline, in seven individuals. Spotty white matter T2 hypersignals were observed in cerebellar, cerebellar peduncles, brainstem and periaqueductal grey matter. Strikingly, PEO was mild or even absent. Muscle biopsy of four affected family members revealed the presence (2-25%) of COX deficient fibers, some of which exhibited 'ragged-red' pathology. Fluorescent imaging of mitochondrial membrane potential in patient fibroblasts revealed a mosaic staining pattern, characteristic of a heterogeneous mitochondrial defect. Long-range and real-time PCR analyses confirmed the presence of clonally-expanded mtDNA deletions as the cause of the focal COX deficiency. Candidate gene screening excluded mutations in POLG, PEO1, RRM2B, SLC25A4 and TK2 whereas sequencing of POLG2 revealed all affected individuals to be heterozygous for a c.970-1G→A splicing mutation. cDNA studies on muscle from two affected individuals indicated 10-15% exon5 skipping and more complex exon5 skipping. By contrast, seven unaffected family members were wild type at this POLG2 locus indicating complete segregation of the clinical phenotype with the novel splice-site mutation. This is the first complex neurological phenotype reported so far due to a mutation in the gene encoding the accessory subunit of human polymerase gamma.
Usher syndrome is the most frequently inherited dual impairment of vision and hearing. Usher syndrome type 1 (USH1) is the most severe form characterized by profound congenital deafness, vestibular dysfunction and prepubescent onset of retinal degeneration. Currently, there are six genes associated with USH1: Myosin 7A (MYO7A), Harmonin (USH1C), Cadherin-23 (CDH23), Protocadherin-15 (PCDH15), Sans (USH1G) and recently identified CIB2. The products of five of these genes have recently been identified to be part of the calsequestrin processes of photoreceptors. The purpose of the study was to genetically characterize a cohort of 38 USH1 probands. Methods: An USH1 cohort of 38, partially pre-screened for MYO7A mutations, was high-throughput sequenced for targeted exons of genes associated with inherited retinal degenerations. Genetically unsolved samples were subsequently screened for deletions using a comparative genomic hybridization (CGH) array and deep intronic mutations were investigated by sequencing of the targeted whole-gene capture of the known Usher genes. Intron regions were mapped onto retina transcriptome. When possible, the likely pathogenic variants were confirmed by co-segregation in available family members with Sanger sequencing. The study protocol adhered to the tenets of the Declaration of Helsinki and was approved by Institutional Review Board. Results: Selective exon capture and Illumina sequencing provided excellent coverage of the targeted exons, with >95% of exons having >10X sequence depth. With this sequencing approach, we were able to solve the Declaration of Helsinki and was approved by Institutional Review Board. Results: Selective exon capture and Illumina sequencing provided excellent coverage of the targeted exons, with >95% of exons having >10X sequence depth. With this sequencing approach, we were able to solve the general genetic cause of the disease, suggesting possible copy number variation in known genes, deep intronic mutations or additional genetic loci for this condition. This work was supported by grants from the National Eye Institute (EY012910) and the Foundation Fighting Blindness.

SPG11 and SPG15 are the most frequent genotypes causing spastic paraplegia with thin corpus callosum, white matter changes and mental retardation in Italian patients. C. Geilera, V. Pensato, B. Castelliotti, E. Sarto, D. DiBella, L. Nanetti, D. Pareyson, E. Salsano, M. Eoli, C. Ciano, C. Mariotti, F. Taroni. 1) Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS - Istituto Neurologico Carlo Besta, Milan, Italy; 2) Unit of Clinic of Central and Peripheral Degenerative Neuropathies, IRCCS - Fondazione Istituto Neurologico Carlo Besta, Milan, Italy; 3) Unit of Molecular Neuro-Oncology, IRCCS - Fondazione Istituto Neurologico Carlo Besta, Milan, Italy; 4) Unit of Clinical Neurophysiology, IRCCS - Fondazione Istituto Neurologico Carlo Besta, Milan, Italy. Hereditary Spastic Paraplegias (HSPs) are clinically and genetically heterogeneous diseases characterized by lower limb spasticity and weakness. The genetic of HSPs is complex and at least 38 HSP genes have been identified as causative ones. A subgroup of complicated autosomal recessive HSPs (ARHSP) has been distinguished for the presence, of thin corpus callosum (TCC) and mild white matter lesions (WML). The ARHSP-TCC phenotype has been recognized in association with at least seven genes but causative mutations have been most frequently found in the KIAA1840 (SPG11) and ZFYVE26 (SPG15) genes. We selected a cohort of 53 unrelated Italian patients with complicated spastic paraplegia presenting with at least one clinical/neuro-radiological hallmark (mental retardation, TCC and/or WML). DNA samples from the patients were analyzed for mutations in SPG11 (KIAA1840), SPG15 (ZFYVE26), SPG21 (ACP33) and SPG5A (CYP7B1) genes. Patients included in this survey presented juvenile onset of neurological symptoms (14.6±10yrs). In 10 cases family history was compatible with a recessive inherited disorder, while the remaining were sporadic cases. Molecular investigations allowed the genetic diagnosis in 21 index cases: 15 patients, from 14 families, were found to carry pathogenic mutations in the SPG11 gene, and 7 unrelated patients were found to carry mutations in the SPG15 gene. The remaining 32 cases were negative also for mutations in SPG21 and SPG5A genes. SPG11 mutations were found in homozygous form in 21 cases and in compound heterozygous form in the remaining 10 patients. Overall, we found 19 different mutations with a large prevalence (74%) of nonsense mutations. Only, 7 mutations have been previously reported. SPG15 mutations were all newly identified mutations. They have been found in homozygous form in 6 cases (five carried a stop mutation and one carried a missense mutation), while one patient was compound heterozygous for a frameshift and a stop mutation. In all cases we could demonstrated the segregation of the identified mutations within families. In our cohort of Italian patients with complicated HSP we found SPG11 mutations in 26% of the cases, and SPG15 mutations in 13%. The majority of mutations are predicted to cause absence of the protein and 18 mutations were newly reported. Negative molecular results in at least 60% of cases, suggests further genetic heterogeneity among the patients with similar clinical presentation.
2971F
5q15 deletions: clinical and molecular characterization of a new syndrome. E. Palleis-Pocachard1, J. Andreux1, A-M. Bisgaard Pedersen2, R. Steensberg Møller2, E. Buhler1, A. Carballo3, P.H. Kaarst1, E. Parrini1, D. Héron1, B. Keren4, B. Benyajita5, N. Sobreira6, V. Malan7, L. Manouvrier-Hanu1, T. Wang9, Y. Guerin1, A. Represa8, M. Kirchhoff1, M. Holder-Espinasse1, C. Cardoso1, 1 Plateforme de Biologie Moléculaire et Cellulaire, INSERM, Marseille, France; 2 Plateforme de génomique, Centre de Biologie Pathologique, CHRU, Lille, France; 3 Department of Clinical Genetics, Rigshospitalet, Denmark; 4 The Willem Johannsen Centre for Functional Genome Research, Department of Molecular and Cellular Medicine, The Panum Institute, University of Copenhagen, Copenhagen, Denmark; 5 INMED, INSERM, Aix-Marseille University, Marseille, France; 6 Department of Paediatrics, Hôpital Universitaire, Hôpitaux de Paris, France; 7 Institute of Genetic Medicine and Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, USA; 10) Département de Cytogénétique et de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 11) Service de Génétiq"e Clinique, Hôpital Jeanne de Flandre, CHRU de Lille, France.

Malformations of Cortical Development (MCD) are important causes of intellectual disability (ID) and account for 20-40% of drug-resistant epilepsy in childhood. A large number of MCD have now been identified and classified using embryology, genetic, and imaging criteria. Contrary to previous assumptions, the majority of these disorders are now thought to have a genetic basis, although environmental causes such as utero infection or ischemia are still possible. Microdeletions of the 5q15-3-g15 region have been described in patients with abnormal brain development, severe ID and epilepsy. Further delineation of a critical region of overlap in these patients pointed to MEF2C as the causative gene. However, we previously reported two patients with a 5q14.3-q15 deletion that did not span MEF2C suggesting that other genes may be implicated in these conditions. Here, we investigated 6 additional patients harboring partially overlapping deletions that do not include MEF2C. The clinical phenotype includes moderate-to-severe mental retardation, microcephaly, abnormal spastic gait, macrodactyly, and seizures. The minimal critical region identified by all patients spans an interval of approximately 4.2 Mb in the 5q15 region. This interval contains 8 genes: NR2F1, FAM172A, POUSF2, CSorf36, ANKRD32, MCTP1, FAM61B and TTC37. Among these genes, we considered three of them as the best candidates since they are expressed in fetal brain tissues. To identify which gene contribute to MCD phenotype, we developed an experimental approach. Indeed, we used RNAi-based inactivation in rodents, enabling rapid functional evaluation of candidate genes potentially impacting neuronal migration and maturation. Using this approach, we succeeded to confirm that one of the candidate gene mapped to 5q15 is crucial for brain development.

2972W
Alteration of ganglioside biosynthesis responsible for complex hereditary spastic paraplegia, SPG26. G. STEVANIN1,2,3, A. Boukhris1, R. Schulte4, C.M. Lourenço5,6,7, A. Ferrero1,2,8, L. Gonzalez9,10, P. ter Brugge11,12,3, P. Charles11,12,3, I. Rekik1, J. Gauthier9, R.F. Acosta-Lebrigo8, F. Spezian1, A. Berbert7, A. Caballero-Oteyza5, A. Dionne-Laporte6, A. Noreau6, M. Gaussen5,7, P. Coutinho5, P. Dion2, M. Fisch2, J. Poulou1,7, C. Mhiiri2, L. Schols5, F. Danos2, G.A. Kroules2, W. Marques Jr5, A. Durr9,11, A. Durr2,9,11, K.Sugai1, M. Kato1,2,3, A. Durr2,9,11, E.Kasai-Yoshida5, N.Sawaura9, M.A. Gonzalez4, C.Mhiri5,12,3, M. Holder-Espinasse1, T. Wang9, Y. Guerin1, A. Represa8, M. Kirchhoff1, M. Holder-Espinasse1, C. Cardoso1, 1) Institute of Genetic Medicine and Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, USA; 10) Département de Cytogénétique et de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 11) Service de Génétiq"e Clinique, Hôpital Jeanne de Flandre, CHRU de Lille, France.

Hereditary spastic paraplegias (HSPs) form a heterogeneous group of neurological disorders. A whole genome linkage mapping effort was made using three HSP families from Spain, Portugal and Tunisia and it allowed us to reduce the SPG26 locus interval from 34 to 9 Mb. Following this, a targeted capture was made to sequence the entire exome of affected individuals from the three above families, as well as from two additional autosomal recessive HSP families of German and Brazilian origins. Five homozygous truncating (n=3) and missense (n=2) mutations were identified in B4GALNT1. Following this finding, we analyzed the entire coding region of this gene in 65 additional cases, and three mutations were identified in two subjects. All mutated cases presented an early onset spastic paraplegia, with frequent intellectual disability, cerebellar ataxia and peripheral neuropathy as well as cortical atrophy and white matter hyperintensities on brain imaging. B4GALNT1 encodes beta-1,4-N-acetyl-galactosaminyl transferase 1 (B4GALNT1), involved in ganglioside biosynthesis. These findings confirm the increasing interest of lipid metabolism in HSPs. Interestingly, while the cataabolism of gangliosides is implicated in a variety of neurological diseases, SPG26 is the second human disease involving defects in their biosynthesis.

Objective: To identify novel mutations causing hereditary motor and sensory neuropathy type V (HMSN-V), a variant of Charcot-Marie-Tooth disease (CMT), we screened known CMT and related genes in members of a Japanese family with HMSN-V. Methods: The clinical features, peripheral nerve conduction properties, and MRI findings of four patients from the Japanese family were examined to confirm HMSN-V. We then screened 28 CMT disease-causing and related genes using a custom microarray chip. Results: The clinical features included mild weakness of the distal lower limb muscles, foot deformity, mild sensory loss, and late-onset progressive spastic paraparesis. Electrophysiological studies revealed slower sensory and motor nerve conduction and small sensory and muscle compound action potentials in multiple nerves in limbs, indicating widespread neuropathy. Brain MRI revealed an abnormally thin corpus callosum. In all four patients, microarrays detected a novel heterozygous missense mutation c.1166A>G (p.Y389C) in the gene encoding the light chain neurofilament protein (NEFL). This mutation is segregated with disease and patient phenotype in this family.

Conclusion: All members of the Japanese family with motor and sensory neuropathy with pyramidal tract involvement harbored the same novel NEFL mutation, indicating that NEFL mutations can result in an HMSN-V phenotype.

2976T KIF5A gene mutations in patients with spastic paraplegia. F. Taroni1, S. Calderazzo1, E. Sirio1, S. Baratta1, K. Savio2, G. Galassi3, A. Anelli4, G. Lauria4, L. Nanetti4, E. Salsano5, D. Pareyson6, C. Mariotti6, D. Di Bella7. 1) Genetics of Neurodegenerative & Metabolic Diseases, IRCCS Ist Neurol Carlo Besta, Milan, Italy; 2) Neurology Unit, Ospedale degli Infermi, Biella, Italy; 3) Neurology Unit, Ospedale di Modena, Modena, Italy; 4) Neuromuscular Disease Unit, IRCCS Ist Neurol Carlo Besta, Milan, Italy; 5) Neurology 8, IRCCS Ist Neurol Carlo Besta, Milan, Italy.

Hereditary spastic paraplegias (HSP) are a clinically and genetically heterogeneous group of neurodegenerative disorders. Pure and complicated forms of the disease have been described. About half of HSP cases result from autosomal dominant (AD) mutations in spastins (SPG4), atlastins-1 (SPG3A), or REEP1 (SPG31) genes. Mutations in the KIF5A gene have been reported to be a relatively frequent cause (approx. 10%) of pure or complicated AD-HSP phenotypes (SPG10) in French and Italian populations. The KIF5A gene encodes the neuronal kinesin heavy chain implicated in the anterograde axonal transport. Most mutations identified so far (all missense except one deletion) are located in the motor domain, with the exception of one mutation identified in the neck region and one in the stalk domain of the protein. The motor domain (exons 1-11) of KIF5A was analysed by high-resolution melting (HRM) analysis and/or direct sequencing in 350 unrelated HSP index cases, including 185 AD and 165 sporadic cases, negative for SPG4 mutations. We identified 10 different missense mutations, 4 of which are novel, in 11 HSP probands. The novel mutations were not detected in >350 normal alleles. By in silico analysis, all mutations are predicted to damage protein structure or function. Six out of seven cases identified in AD-HSP probands, 3 cases were apparently sporadic while for 2 probands family history was either not available or unclear. Our patients showed a wide range of age-at-onset, with 4 patients exhibiting onset in childhood, 2 patients at >50 yrs, and the remaining patients in the 3rd-4th decade. Interestingly, 9/11 patients had a complex phenotype with axonal neuropathy, parkinsonism, ataxia, and deafness as variably associated symptoms. Conclusions: In our study, KIF5A mutations account for 3.5% of cases with autosomal dominant inheritance or complicated AD-HSP. The frequency lower than that observed in other Italian (Crimella 2012) and French (Goizet 2009) patient series but similar to that (4%) found in a mixed European population (Schüle 2008). Our findings indicate that SPG10 should be considered in the molecular analysis of HSP phenotypes, even in the absence of a clearly positive family history. Clinically, the probands of our 28 SPG10 families present with a rather broad range of phenotypes, with polyneuropathy or Charcot-Marie-Tooth phenotype as the most frequently associated feature. (Partly supported by grant RF-2009-1539841 from Italian Ministry of Health to F.T.).

2977F A Locus for Autosomal Recessive Congenital Mirror Movement Disorder Maps to 2q23.1. J.B. Vincent1, M.A. Raffa1, P. John1, J.A. Balouch2. 1) Neurogenetics Sect, R30, Ctr Add/Mental Hlth,Clarke Div, Toronto, ON, Canada; 2) Alta-ur-Rehan School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad-Pakistan.

Mirror Movement (MRRM) disorder is a rare, mainly autosomal disorder, although it is thought that some sporadic cases may be due to recessive inheritance. MRRM refers to voluntary movements on one side of the body being mirrored by involuntary movements on the opposite side. Using a linkage analysis and candidate gene approach two genes so far has been implicated in MRRM: DCC on 18q21.2, which encodes a netrin receptor, and on 15q15.1, which is involved in the maintenance of genomic integrity. Here, we describe a large consanguineous Pakistani family with cases 11 of MRRM reported across 5 generations, with autosomal recessive inheritance likely. We used Sanger sequencing to exclude DCC and RAD51. We then employed microarray genotyping and autozosity mapping to identify a shared region of homozygosity-by-descent among the affected individuals. We identified a single autozygous region of ~3.3 Mb on chromosome 22q13.1 (Chr22:36605976-39904648). We used Sanger sequencing to exclude several candidate genes within this region, including DM1. Whole exome sequencing is currently underway to identify the causative mutation in this family.

2978W The Whole Genome Sequence of a Jack Russell Terrier with Progressive Spinocerebellar Ataxia and Myokymia Contains a Homozygous Disease-Associated KCNJ10 Missense Mutation. D. Giliam1, J.R. Coates2, G.S. Johnson1, L. Hansen3, Y. Mihanga-Mutangadura1, J.F. Taylor3, G.G. Johnson2, R.D. Schnabel3, D.P. O’Brien2. 1) Department of Veterinary Pathobiology, University of Missouri College of Veterinary Medicine, Columbia, MO; 2) Department of Veterinary Medicine and Surgery, University of Missouri College of Veterinary Medicine, Columbia, MO, USA; 3) Division of Animal Sciences, University of Missouri College of Agriculture and Natural Resources, Columbia, MO, USA.

A distinct progressive autosomal recessive spinocerebellar ataxia occurs in three related dog breeds: Jack Russell Terriers, Parson Russell Terriers, and Russell Terriers (JRT/PR/RTs). Affected dogs develop cerebellar ataxia by 2-6 months of age. As the disease progresses, most dogs exhibit myokymia and some have seizures. We used Illumina sequencing technology to generate a whole genome sequence (WGS) with 20-fold coverage from an affected JRT and aligned the reads to the canine reference genome sequence using NextGENE software. The detected sequence variants were prioritized according to predicted functionality and filtered to remove variants also present in any of 22 WGSs generated for dogs with other heritable diseases which served as controls. Among the 107 homoygo- sese sequence variants in 89 genes that survived the filter step, a KCNJ10:c.627C>G transversion predicting a p.I209M amino acid substitution was considered most likely to be causal because cerebellar ataxia and seizures are features of SeSAME syndrome caused by mutations in human KCNJ10. In addition, p.209I and surrounding amino acids are highly conserved components of the functionally important inward rectifier potassium channel domain. We used a TaqMan allelic discrimination assay to determine the genotypes of individual dogs at KCNJ10:c.627. The 13 JRT/PR/RTs in our collection with spinocerebellar ataxia were all homozygous for the KCNJ10:c.627G allele; whereas, only one of the 879 genotyped JRT/PR/RTs with undetermined phenotype was a KCNJ10:c.627G homozygote. We were unable to ascertain any clinical information for this KCNJ10:c.627G homozygote. The other genotyped JRT/PR/RTs were either heterozygotes (n = 100) or homozygous for the reference allele (n = 779). One hundred and two dogs from 62 other breeds were all KCNJ10:c.627G homozygotes. There appear to be species differences between human and canine KCNJ10 deficiency diseases, SeSAME syndrome patients are deaf; whereas, the affected JRT/PR/RTs have normal hearing. In addition, although myokymia is often exhibited by affected JRT/PR/RTs, it is not considered a feature of SeSAME syndrome. This report illustrates that the mutation identified for an unmapped disease can be identified from the WGS of a single affected dog plus controls.
2979T
The Alu-rich architecture of SPAST predisposes to diverse and functionally distinct CNV alleles. P.M. Boone, 1, I.M. Campbell, 1, B.C. Baggett, 1, J.C. Schull, 1, C.J. Shaw, 1, M.A. Withers, 1, P. Moretta, 1, J.K. Fink 2, A. Ordoñez-Ugalde 2, B. Quintans 1, M.-J. Sobrido 3, S. Stemmmer 1. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Medical Genetics Laboratories, Baylor College of Medicine; 3) Dept. of Neurology, Baylor College of Medicine; 4) Michael E. DeBakey VA Medical Center, Houston, TX, USA; 5) Children's Medical Center of Dallas, TX, USA; 6) Department of Neurology, University of Michigan, Ann Arbor, MI, USA; 7) Grupo de Medicina Xénómica, Fundación Pública para la Ciencia y la Salud, Xénónicas, S.L., Avda. de la Constitución 12, 38012, Girona, Spain; 8) Department of Human Genetics, Ruhr University Bochum, Germany; 9) Texas Children’s Hospital, Houston, TX, USA; 10) Department of Pediatrics, Baylor College of Medicine.

SPAST is the disease gene for autosomal dominantly inherited spastic paraplegia type 4 (SPG4). While disease-causing deletions and duplications (CNVs) in SPAST have been described, their origin and molecular consequences remain obscure. We mapped 44 SPAST CNVs with nucleotide resolution in subjects with SPG4. A diverse combination of exons were deleted or duplicated, providing insights into the importance of particular exons for spastin function. 75% of CNV breakpoints localized to Alu elements, suggesting that the Alu-rich architecture of SPAST renders this locus susceptible for altered production or altered localization of key factors in autophagy, and glaucoma.

2979W
Lamin B1 over-expression is associated with nuclear defects and alterations in nuclear structure. H. Rolyan 1, K. Dahli 2, N. Dahli 2, A. Melberg 2, C.J. Smith 2, C. Combes 1, M. Vairo 3, A. Brusco 3, Q.S. Radiath 1. 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, USA; 3) Dept. of Immunology, Genetics and Pathology, Section of Clinical Genetics, The Rudolphi Laboratory, Uppsala University Children’s Hospital, Uppsala, Sweden; 4) Department of Neuroscience, Neurology, Uppsala University, Uppsala Sweden; 5) Hospital de Clinicas de Porto Alegre & Universidade Federal do Rio Grande do Sul; Brazil; 6) University of Torino, Department of Medical Sciences, Torino, Italy.

Autosomal Dominant Leukodystrophy (ADLD) is an adult onset demyelinating disease caused by duplications of the lamin B1 (LMNB1) gene. LMNB1 is a component of the nuclear lamina and together with providing structural integrity to the nucleus, it also plays an important role in gene regulation and chromatin modification. The mechanism by which LMNB1 duplications cause ADLD is unknown and it is also not known whether LMNB1 duplications lead to altered nuclear structure. To address this question, we have analyzed the nuclear structure in fibroblasts from ADLD patients. We have shown for the first time that patient fibroblasts show a greater number of nuclear abnormalities, alterations in nuclear shape and size and have altered chromatin laminar architecture. These results were confirmed in mouse fibroblast lines over expressing LMNB1. Our results suggest that alterations in nuclear structure are a functional consequence of LMNB1 over expression that may play an important role in disease causation.

2982T
Mechanisms underlying non-recurrent microdeletions in neurofibromatosis type-1 (NF1). J. Vogt 1, K. Bengesser 1, K. Claes 1, K. Wimmer 1, L. Messiien 2, V.-F. Mauthner 3, R. van Minkelen 1, E. Legius 1, H. Brems 1, T. Rosenbaum 1, M. Upadhyaya 4, D.N. Cooper 5, H. Kehrer-Sawatzky 6, 1) Institute of Human Genetics, Ulm University, Ulm, Baden-Württemberg, Germany; 2) Center for Medical Genetics, Ghent University Hospital, 9000 Ghent, Belgium; 3) Division of Human Genetics, Medical University Innsbruck, Innsbruck, Austria; 4) Medical Genomics Laboratory, Department of Genetics, University of Alabama at Birmingham, Birmingham, USA; 5) Department of Neurology, University Hospital Hamburg-Eppendorf, Hamburg, Germany; 6) Department of Clinical Genetics, Erasmus MC, Rotterdam, Netherlands; 7) Department of Human Genetics, KU Leuven, Leuven, Belgium; 8) Clinical Centre Duisburg, Children’s Hospital, Germany; 9) Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK.

NF1 microdeletions encompassing the NF1 gene region at 17q11.2 are present in ~5% of patients with NF1. Although the mechanisms underlying recurrent NF1 microdeletions are well characterized, those underlying non-recurrent (atypical) NF1 microdeletions are not well delineated. NF1 microdeletions with non-recurrent breakpoints are heterogeneous in terms of their size, breakpoint position and number of deleted genes. In this study, we have analyzed a cohort of atypical NF1 deletions using high-resolution custom aCGH and performed breakpoint-spanning PCRs with primers located in non-deleted regions closely flanking the deletion boundaries. Sequence analysis of breakpoint-spanning PCR products indicated that 15% of atypical NF1 deletions contain breakpoints located within introns and hence were probably mediated by NHEJ/MMEJ. We therefore conclude that NHEJ/MMEJ is the most common mechanism underlying atypical NF1 deletions. However, two of the 20 deletions investigated exhibited complex breakpoint junctions without further complexity. These deletions were analyzed, as determined by inverse PCR, semi-specific PCR and Genome Walker analysis. Large genomic deletions associated with SVA-element insertions at the breakpoints have not been previously reported. We postulate that SVA elements, transcribed during retrotransposition, that trigger genomic rearrangements may represent an as yet uncharacterized mechanism responsible for causing CNVs with non-recurrent breakpoints. We noted that 14 (70%) of the 20 atypical NF1 microdeletions displayed proximal breakpoints located within 38-kb region of the SUZ12P. The enrichment of atypical NF1 deletion breakpoints in SUZ12P is remarkable since the breakpoints of recurrent type-2 NF1 deletions are also located within SUZ12P. However, only two atypical NF1 deletion breakpoints co-localized with the breakpoints of previously analysed type-2 NF1 deletions. Hence, a clustering of type-2 SUZ12P breakpoints in the region of the atypical NF1 deletion breakpoints was not observed.

Nevertheless, the accumulation of breakpoints associated with recurrent and non-recurrent NF1 deletions within SUZ12P is indicative of its genomic instability. The analysis of FISH, microsatellite markers and insertion/deletion site analysis confirmed that atypical NF1 deletion patients investigated. Thus, not only recurrent type-2 NF1 deletions, but also a considerable proportion of atypical NF1 deletions, are of postzygoty origin.
We establish a new link between alteration of mitochondrial protein synthesis and West syndrome (WS) since many of the causative nuclear genes are probably yet unknown. This heterogeneity often impedes the identification of the etiology of WS within the clinical practice. We identified using exome sequencing a homozygous mutation in the GUF1 gene in three siblings from a consanguineous family affected with isolated West syndrome and severe neurological impairment. The encoded GUF1 protein is a universally conserved GTPase in mitochondria and chloroplasts. Its prokaryote homolog, named EF4 or LepA, encodes a non-essential elongation factor both in eu- and prokaryote could explain its ubiquitous distribution. In the two patient groups, we identified and which segregates with WS modifies an Arginine residue conserved in all eukaryotic organisms from budding yeast to human. It is positioned within the beta-strand moiety of the GTPase domain that specifically recognizes the acceptor stem and D-loop of tRNA molecules. Such defects affecting mitochondrial protein synthesis might lead to severe epileptic disorders in infants such as IS/WS. They are, however, easily overlooked possibly because of the difficulty to correctly gauge the activity of the respiratory chain complexes, especially when no lactate and/or pyruvate increase is identified in blood and/or cerebrospinal fluid. We show that exome sequencing is a suitable strategy for identifying causative genes in familial IS/WS since many of the causative nuclear genes are probably yet unknown. We establish a new link between alteration of mitochondrial protein synthesis and IS/WS.
2986F
HIPPOCAMPAL DEFECTS IN THE FMR1 KNOCKOUT MOUSE. F. Kooy1, I. Heulens2, V. Sabanov3, T. Ahmed2, A. Popp2, R. Willemsen2, R. D’Hooghe2, D. Baitschun2, L. Rooms1; 1) Dept Medical Genetics, Univ Antwerp, Antwerp, Belgium; 2) Laboratory of Biological Psychology, KU Leuven, Belgium; 3) Department of Clinical Genetics, Erasmus MC Rotterdam, The Netherlands.

Whole-cell patch-clamp recordings on hippocampal slices from CA1 pyramidal neurons revealed a significant decrease (p=0.0127) of IPSCs in three-week-old Fmr1 knockout mice. Several drugs that modulate the GABA(A) receptor are available for clinical use. Ganaxolone, a positive allosteric modulator of GABA(A) receptors, prevents seizures in Fmr1 knockout mice. To further explore the therapeutic effect of ganaxolone, we evaluated the effect of the drug in the marble burying paradigm. Acute ganaxolone treatment is able to correct the abnormal marble burying behavior of Fmr1 knockout mice without having an effect in wild-type littermates (p<0.001). Previous studies reported enhanced mGlur5-dependent long-term depression (LTD) in Fmr1 knockout mice. We examined NMDA-dependent LTD by induction with low-frequency stimulation. Our data reveal a significant decrease in this form of LTD in Fmr1 knockout mice (p<0.05). Previously, we found a reduced expression of Larg in Fmr1 knockout mice. LARG is a GEF which activates RhoA, involved in actin cytoskeleton remodeling and spine morphology. We show that several genes of the LARG-RhoA pathway are underexpressed in the fragile X syndrome. DiOlistic labeling shows that spine density was significantly increased in the CA1 region of the hippocampus in Larg−/− mice (p<0.01), while spine length remained unchanged. Underexpression of LARG thus results in dendritic spine abnormalities with striking similarities to those observed in the fragile X syndrome.

2986T
ComprehensiveMutationAnalysisforHereditarySensoryandAutonomicNeuropathyUsingaNext-GenerationSequencingSystem. J. Yuán1, E. Matsura2, Y. Higuchi3, A. Hashiguchi4, T. Nakamura5, S. Nozuma2, Y. Sakiyama1, A. Yoshimura6, H. Takashima1; 1) Kagoshima University School of Medicine, Kagoshima, Japan; 2) 2987W
Analysis of Genomic DNA sequence based on new molecular diagnostic strategies for Neurofibromatosis Type I. S. Choi1, J. Kim1, J. Lee2, G. Kim3, B. Lee4, H. Yoo5; 1) Medical Genetics Center, Asan Medical center, Seoul, South Korea; 2) Dept. Pediatrics Asan Medical Center, Children’s Hospital, University of Ulsan College of Medicine, Seoul, South Korea.

Neurofibromatosis Type I (NF1, OMIM 162200) is characterized by multiple café au lait spots, axillary and inguinal freckling, multiple cutaneous neurofibromas, and iris Lisch nodules. NF1 is diagnosed based on the constellation of clinical findings. The genetic testing in NF1 is also important for the diagnostic confirmation. More importantly, it is essential for the appropriate genetic counseling. However, the molecular genetic testing is problematic due to the large size of the causative gene, NF1. In addition, the gene has homologous domains such as Cystein-Serine rich domain (CSD), GTPase related domain (GRD), and adenylated kinase pseudogene (AK3). These domains are dispersed in multiple chromosomal loci, leading to false positive results in routine medical sequencing tests in NF1. To avoid interferences by these homologous sequences, labor-intensive RNA-based sequencing methods have been used. To facilitate the genetic diagnostic process with high accuracy, we introduced the new strategy of genomic DNA-based molecular test for NF1. A total of eighteen patients were enrolled. All patients were diagnosed with NF1 on the basis of typical clinical features. Functional studies of the identified variants and genes will be identified.

2987W
Analysis of Genomic DNA sequence based on new molecular diagnostic strategies for Neurofibromatosis Type I. S. Choi1, J. Kim1, J. Lee2, G. Kim3, B. Lee4, H. Yoo5; 1) Medical Genetics Center, Asan Medical center, Seoul, South Korea; 2) Dept. Pediatrics Asan Medical Center, Children’s Hospital, University of Ulsan College of Medicine, Seoul, South Korea.

Neurofibromatosis Type I (NF1, OMIM 162200) is characterized by multiple café au lait spots, axillary and inguinal freckling, multiple cutaneous neurofibromas, and iris Lisch nodules. NF1 is diagnosed based on the constellation of clinical findings. The genetic testing in NF1 is also important for the diagnostic confirmation. More importantly, it is essential for the appropriate genetic counseling. However, the molecular genetic testing is problematic due to the large size of the causative gene, NF1. In addition, the gene has homologous domains such as Cystein-Serine rich domain (CSD), GTPase related domain (GRD), and adenylated kinase pseudogene (AK3). These domains are dispersed in multiple chromosomal loci, leading to false positive results in routine medical sequencing tests in NF1. To avoid interferences by these homologous sequences, labor-intensive RNA-based sequencing methods have been used. To facilitate the genetic diagnostic process with high accuracy, we introduced the new strategy of genomic DNA-based molecular test for NF1. A total of eighteen patients were enrolled. All patients were diagnosed with NF1 on the basis of typical clinical features. Functional studies of the identified variants and genes will be identified.
2990W  
Investigating the genetic basis of amyotrophic lateral sclerosis using next-gen techniques. J.A. Filitta1, K.L. Williams1, G.A. Nicholson2, J.P. Blair1, 1) Macquarie University, Sydney, New South Wales, Australia; 2) Molecular Medicine Laboratory, Concord Hospital, New South Wales, Australia.

Amyotrophic lateral sclerosis (ALS) is a fatal form of motor neuron disease characterized by the progressive degeneration of both the upper and lower motor neurons. ALS may occur as familial or sporadic disease. To date, genetic mutations are the only known cause of ALS. ALS is heterogeneous with mutations in over 15 genes now described in both familial and sporadic ALS, including SOD1, TARDBP, FUS, C9ORF72, UBNL2, OPTN, PFN1, hnRNPA1 and SS18L1. Nevertheless, the genes are yet to be identified in 30-40% of familial ALS. We aim to identify novel genes involved in ALS using next-generation sequencing techniques. We performed exome capture (Illumina TrueSeq Exome Enrichment Kit) and massively parallel sequencing (Illumina HiSeq2000) on 14 DNA samples from five ALS families that are negative for all known ALS genes. The families used in this study have DNA samples available from either two affected individuals and a married-in control, or more than two affected individuals. Bioinformatic analysis of exome data identified several candidate variants among four families. Additional segregation analysis in one of these families reduced the number of potential causal variants to two. Analysis of large control cohorts is underway to validate these variants. Functional studies will then be undertaken to examine the pathological consequences of these variants in an attempt to further understand the disease mechanisms underlying both sporadic and familial ALS. Using a combination of exome sequencing and Sanger sequencing we have also determined the prevalence of known ALS genes in Australian familial and sporadic ALS cohorts. Mutations in known ALS genes account for 57.2% of familial ALS and 5% of sporadic ALS. The familial mutations comprise of SOD1 (13.9%, n=26), TARDBP (2.1%, n=4), FUS (2.7%, n=5), C9ORF72 (38%, n=71), OPTN (0.5%, n=1), UBNL2 (1.1%, n=2), and SS18L1 (0.5%, n=1). Currently, the cause of 40% of ALS families, and 95% of sporadic ALS is still unknown. Additional exome sequencing projects are underway in Australian familial index patients to identify novel ALS genes.

2991T  
Molecular genetics of primary microcephaly in Indian population: mutations in WDR62, ASPM and STIL genes. A. Kumar1, V. Bhat1, G. Mohan1, S.C. Girnijati2. 1) Molecular Reproduction, Development & Genetics, Indian Institute of Science, Bangalore, Karnataka, India; 2) Child and Adolescent Psychiatry, National Institute of Mental Health and Neuro Sciences, Bangalore, Karnataka, India.

Development of the human brain, specially the cerebral cortex, is an intricate and complex process that involves generation of neural progenitors in periventricular zones, cell proliferation through symmetric and asymmetric cell divisions, and finally the migration of post-mitotic neurons to their final destinations. An understanding of the molecular mechanisms guiding these intricate interdependent processes is facilitated by the discovery of causative genes for a Mendelian form of the cortical (brain) malformation disorder, primary microcephaly (MCPH), which is characterized by a smaller than normal brain and mental retardation. Interestingly, the brain volume (400 cc) of an MCPH patient is similar to early hominids, such as Australopithecus who lived 2.0-4.4 myr ago. Occasionally, MCPH patients also have other brain malformations, such as pachygyria, polymicrogyria, simplified gyral pattern etc. We have been working on the genetics of MCPH for some time, having identified three novel MCPH loci in Indian families, each with two affected individuals with primary microcephaly. We have been able to discover the causative genotypes in most of these families and found a role in the coordination of mitotic processes as well as mitotic spindle formation.

2992F  

Introduction: Hearing impairment is the most common sensory disorder in humans, affecting approximately one to three in 1000 new borns, with 50% due to genetic causes. The majority of these cases (70%) are nonsyndromic, about 2% of these are X-linked. So far five different X-linked loci have been mapped, the causative gene POU3F4 (MIM 300039) has been identified for the gene locus DFNX2. This hearing loss is often progressive, with temporal bone abnormalities and stapes fixation. Temporal bone abnormalities include dilatation of the fundus of the internal acoustic canal (IAC), that can increase the risk of CSF gusher during cochlear implant surgery. During inner ear development, the homologous mouse Pou3f4 gene is exclusively detected in the mesenchymal tissue adjacent to the otic epithe-

2993W  
Using whole exome sequencing to diagnose primary microcephaly caused by mutations in ASPM and WDR62. A.H. Cobi1, D. Pehlivan2, E. Karaca2, M.M. Atik3, T. Tos3, W. Wizniewski2, Y. Bayram2, D. Muzny4, R.A. Gibbs4, J.R. Lupski2,4,5. 1) Medical Genetics, Karadeniz Technical University, Faculty of Medicine, Trabzon, Turkey; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 3) Department of Medical Genetics, Sami Ulus Children’s Hospital, Ankara, Turkey; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 6) Texas Children’s Hospital, Houston, TX, USA.

Whole-exome sequencing (WES) is a new groundbreaking technology that has changed the course of human genetics. As a more efficient and cost effective method, WES is now widely used as a diagnostic tool for identifying genetic disorders. Microcephaly is defined by an occipito-frontal circumference less than 2 standard deviations (SD) under age and gender-based averages. Mutations in the ASPM and WDR62 genes have been found to cause primary microcephaly. The ASPM gene probably plays a role in the coordination of mitotic processes as well as mitotic spindle regulation and may also have a function in regulating neurogenesis, whereas the WDR62 gene codes for a spindle pole protein that serves in cerebral cortical development. Here, we present two separately consanguineous families, each with two affected individuals with primary microcephaly. We identified homozygous frameshift deletion in ASPM gene in the affected individuals from one family and WDR62 gene in the other family, whereas parents were at heterozygous state for this mutation.
2994T
ZBTB20-dependent transcription regulation imbalance in autism and intellectual disability. R. Koul1,2, L. Dukes-Rimsky3, Y. Luo1,2, K.A. Jones3, E. Lemyre3, S. Swoll3, D.P. Sprague3, S. Luedde3, B.R. DuPont1, L.C.M. Wilson3, C. Skinner3, F. Gurrieri4, R.E. Stevenson4, E. Boyd4, J.L. Michaud5, L. Wang6, P. Penzes3,4, A.K. Srivastava1,2, J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC; 2) Department of Genetics and Biochemistry, Clemson University, Clemson, SC; 3) Department of Physiology, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL; 5) CHU Sainte-Justine, Centre de recherche, Montréal (QC), Canada; 6) Institute of Medical Genetics, Catholic University, Rome, Italy. 8) Present address: Department of Pediatrics, Emory University School of Medicine, Atlanta, GA.

Dendritic spine morphology and dendritic arborization are key determinants of neuronal connectivity, and have been found to be altered in many neurodevelopmental disorders such as autism spectrum disorders (ASDs) and intellectual disability (ID). Several studies suggest that the changes in neuronal gene expression controlled by selective expression of transcription factors (TFs) affect the formation of dendritic spines and synapses. Recently, we identified the disruption of two transcription factor genes, ZBTB20, a member of the BTB/POZ (poxvirus and zinc finger) family of transcription factors, and SOX5, a member of the SOX (SRY-related HMG-box) family of transcription factors, by a chromosome translocation (t(3;12)) in a patient with developmental delay and autistic features. In addition, we identified two SOX5 mutations, one missense and one nonsense, unique to non-syndromic ID, two non-synonymous ZBTB20 mutations unique to ASD and to ASD and ID, and a de novo intragenic deletion of ZBTB20 in a patient with borderline intelligence, attention deficit and impulsivity. Recently, several genomic de novo mutations involving either ZBTB20 or SOX5 have also been reported in patients with ASD and/or ID. Studies in mice brain have shown that ZBTB20 functions as a transcriptional repressor that binds and represses several TF genes, including SOX5, which are involved in neurogenesis, dendritogenesis, and neuronal circuit formation. We found that in human brain, ZBTB20 expression was largely negatively correlated with these co-expression TF genes and genes enriched with Gene Ontology terms involving synapse and synaptic transmission functions. In cultured rat cortical pyramidal neurons, overexpression of ZBTB20 affected spine morphology and dendritic arborization. Furthermore, overexpression of ZBTB20 in HEK293T cells resulted in significantly decreased expression of TF genes such as TBR1 and MECP2, previously associated with ASD and ID. The ASD/ID-associated ZBTB20 mutants had no repressor activities and caused no change in the expression of TBR1 and MECP2. Taken together, our study suggests the association of SOX5 and its transcription regulator ZBTB20 with ASD and ID and provides potential molecular and cellular mechanisms for ZBTB20 physiological actions, and suggests a potential contribution of a ZBTB20-dependent transcription regulation mechanism in neurocognitive and behavioral disorders.

2996W

Mitochondrial disorders are genetically heterogeneous group of diseases, involving approximately 1,500 genes encoded in the nuclear genome and approximately 1500 genes encoded in the nuclear genome. Mitochondrial myopathy with episodic hyper-creatine kinase-emia (MIMECK) is characterized by episodic or persistent muscle weakness and elevated serum CK levels triggered by infections, drugs, or stressful situations. We previously reported a similar phenotype with mitochondrial DNA (mtDNA) alterations. However, the pathophysiological mechanism of MIMECK is still unknown. The development of Next Generation Sequencing (NGS) has revolutionized the diagnostic approach. New approaches make genetic analyses much faster, more sensitive and more efficient. The MiSeq personal sequencing system was performed on 30 samples and revealed mtDNA deletions in muscles samples of genetically proven cases, clinically and pathologically suspected cases and MIMECK. We found low rate mtDNA deletions in MIMECK and also found various rate mtDNA deletions heteroplasmy in other mitochondrial disease including MELAS with 3243A>G mutation. Further study we need to perform analysis for nuclear gene or pathological study to detect pathophysiology. Sanger sequencing does not have the sensitivity to detect heteroplasmy mutations below about 20%. It also does not detect large deletions. Southern blot analysis is not sensitive to detect low levels of heteroplasmy deletions. Low rate mtDNA deletions analysis by NGS could be more sensitive tool before exome-sequence for nuclear mitochondrial related gene or underdiagnosed mitochondrial disease.

2997T
Severe Congenital RYR1 Associated Myopathy: AR and AD RYR1 Mutations that Expand the Genetic, Clinical and Pathological Spectrum. L. Medne1, D.X. Bharucha-Goebel1, M. Santh1, Z. Zukosky2, J. Dastgir3, T. Winden4, P.B. Shieh5, G. Tenenkeoon6, R.S. Finkel7, J.J. Dowling8, N. Monnier9, C.G. Bönnemann7. 1) Div of Neurology, CHOP, Philadelphia, PA; 19104; 2) Dept of Pathology & Laboratory Medicine, CHOP, Philadelphia, PA 19104; 3) Neurogenetics Branch, NINDA, NIH, Bethesda, MD; 4) Prevention Genetics, Marshfield, WI; 5) Dept of Neurology, University of California, Los Angeles, CA; 6) Division of Neurology, Nemours Children’s Hospital, Orlando, FL; 7) Department of Pediatrics, University of Michigan Medical Center, Ann Arbor, MI; 8) Biochimie et Genetique Moleculaire, Institut de Biologie et Pathologie, CHU, Grenoble, France.

RYR1 gene mutations are known to cause central core or minocore myopathy, typically with a static or slowly progressive course, and pharmacogenetic predisposition to malignant hyperthermia (MH). We report a series of 13 patients with early neonatal onset RYR1 associated myopathy due to both autosomal recessive and dominant acting RYR1 mutations representing the severe end of the clinical spectrum, thus expanding the genetic, clinical and pathological spectrum of RYR1 associated myopathy. Clinical features associated with the severe neonatal presentation of RYR1 associated myopathy included decreased fetal movement, hypotonia, poor feeding, respiratory involvement, arthrogryposis, ophthalmoplegia in four patients, and femur fractures or hip dislocation at birth. One patient had a cleft palate, and another a congenital rigid spine phenotype. Muscle biopsy was performed in all patients at ages ranging from 3 weeks of life to 4 years old. Marked type I fiber predominance approaching type I uniformity was common. All demonstrated diffuse or extensive central cores, with or without peripheral rimmed artifact. Both recessive and dominant mutations were identified in RYR1. Mutations (de novo or due to mosaic parent) were present in 4 patients and 9 patients had recessive RYR1 mutations. Two of the carrier parents showed muscle weakness and one had ophthalmoplegia, thus showing that those mutations can be dominant acting. This series confirms and expands the clinical and pathological variability associated with both recessive and congenital RYR1 associated myopathy. Both dominant and recessive mutations of the RYR1 gene can result in a severe neonatal onset phenotype, but more clinical and histological heterogeneity has been seen in those with recessive mutations. Absence of cores on muscle biopsy should not preclude the consideration of RYR1 associated myopathy. Furthermore, identification of predominantly recessive RYR1 mutations with severe neonatal phenotype alters recurrence risk counseling for parents and their own maternal and paternal grandparents. Clinical features in heterogeneous carrier parents and siblings given the AD nature of isolated MH predisposition mutations.
2998F
Subtle mutations in the SMN1 gene in Chinese patients with SMA: The p.Arg288Met mutation causing SMN1 transcript exclusion of exon 7. Y. QU1, J. DU1, E. LI1, J. BAI1, Y. JIN1, H. WANG1, F. SONG1, 1) Capital Institute of Pediatrics, Beijing, China; 2) Children’s Hospital Affiliated Capital Institute of Pediatrics, Beijing, China.

Background: Proximal spinal muscular atrophy (SMA) is a common neuromuscular disorder resulting in death during childhood. Around 81 – 95% of SMA cases are a result of homozygous deletions of survival motor neuron gene 1 (SMN1) gene or gene conversions from SMN1 to SMN2. Less than 5% of cases showed rare subtle mutations in SMN1. Our aim was to identify subtle mutations in Chinese SMA patients carrying a single SMN1 copy. Methods: We examined 14 patients from 13 unrelated families. Multiplex ligation-dependent probe amplification analysis was carried out to determine the copy numbers of SMN1 and SMN2. Reverse transcription polymerase chain reaction (RT-PCR) and clone sequencing were used to detect subtle mutations in SMN1. RT-PCR results were confirmed by direct sequencing. Results: Six subtle mutations (p.Ser8LysX23, p.Glu134Lys, p.Leu228X, p.Ser230Leu, p.Tyr277Cys, and p.Arg288Met) were identified in 12 patients. The p.Tyr277Cys mutation was not reported previously. The p.Ser8LysX23, p.Leu228X, and p.Tyr277Cys mutations have only been reported in Chinese SMA patients and the first two mutations seem to be the common ones. Levels of full length SMN1 (f-SMN1) transcripts were very low in patients carrying p.Ser8LysX23, p.Leu228X or p.Arg288Met compared with healthy carriers. In patients carrying p.Glu134Lys or p.Ser230Leu, levels of f-SMN1 transcripts were reduced but not significant. The SMN1 transcript almost skipped exon 7 entirely in patients with the p.Arg288Met mutation. Conclusions: Our study reveals a distinct spectrum of subtle mutations in SMN1 of Chinese SMA patients from that of other ethnicities. The p.Arg288Met missense mutation possibly influences the correct splicing of exon 7 in SMN1. Mutation analysis of the SMN1 gene may provide further insights into the pathogenesis of SMA.

2999W
Exome sequencing identifies PINCH2 mutations associated with early-onset autosomal recessive LGMD with severe cardiomyopathy and triangular tongue. J. Warman Chardon1, A. Smith, J. Woulfe1, K. Rakha1, C. Dennie1, J. Schwartzentruber1, C. Beaulieu1, 1) Capital Institute of Pediatrics, Beijing, China; 2) Children’s Hospital Affiliated Capital Institute of Pediatrics, Beijing, China.

Limb Girdle Muscular Dystrophy (LGMD) is a heterogeneous group of inherited disorders leading to progressive muscle degeneration often associated with cardiac complications. We investigated affected siblings with a childhood onset LGMD with macroGLOSSIA and calf enlargement. The siblings developed decreased ejection fraction with global left ventricular dysfunction in their 3rd decade, severe quadruparesis and relative sparing of the face, with a bizarre triangular tongue. After negative genetic evaluation for known LGMD genes, we performed whole exome sequencing of the affected siblings and identified shared compound heterozygous missense mutations in exon 5 (c.C356T; p.Pro119Leu and c.C342G; p.Asns14Lys) and exon 11 (c.T1034C; p.Leu345Pro) of the PINCH2 gene, which segregated appropriately in the family. Both novel variants in exon 5 are in cis. PINCH2 is an important member of the IPP (ILK, Parvin, PINCH) heterotrimeric complex and is essential for signaling through integrin adhesion receptors that regulate cell migration, spreading and adhesion and is found in skeletal and cardiac muscle cells. The IPP complex stabilizes expression of the component proteins by reducing proteasomal degradation. PINCH1 and PINCH2 demonstrate overlapping expression, competitive binding to ILK and functional redundancy. There is evidence in murine models that PINCH1 may compensate for loss of PINCH2 in some tissues. Suggesting that a potential therapeutic approach may involve upregulation of PINCH1 to stabilize the IPP complex in patients with PINCH2 mutations. Although well demonstrated in animal models, this is a novel discovery of IPP complex dysfunction implicated in LGMD.

3000T
C19orf12 mutations in Neurodegeneration with Brain Iron Accumulation (NBIA). G. Annessa1, P. Tarantino1, M. Gagliardi1, G. Lesca1,2, E. Brousse1,2, A. Gambardella1, A. Quattrone1, 1) Inst. of Neurological Sciences, National Research Council, Mangone (CS), CS, Italy; 2) Department of Medical Genetics, Hospice Civiles de Lyon and ClaudeBernard Lyon, University Lyon France, Lyon, France.; 3) CRNL, CNRS UMR 5292, INSERM U1028, Lyon, France Lyon, France; 4) Université Lyon I; Centre de Neurosciences Cognitives, CNRS UMR5292, Lyon, France.; 5) Service de Neurologie C. Hôpital Neurologique Pierre Wertheimer, Lyon, France, Lyon, France.; 6) Neuroimaging Research Unit, National Research Council, Catanzaro, Italy, Catanzaro, Italy.

Neurodegeneration with Brain Iron Accumulation (NBIA) comprise heterogeneous group of progressive neurodegenerative disorders that present with a progressive extrapyramidal syndrome and excessive iron deposition in the brain, particularly affecting the basal ganglia, mainly the globus pallidus. Genetic defects in PANK2 gene are the most common cause of NBIA, followed by mutations in PLA2G6, few reported NBIA families are known to carry mutations in FAH2, a gene previously associated with familial leukodystrophy and spastic paraparesis. Mutations within C19orf12 have recently been identified in patients with NBIA. This gene C19orf12 codes for a mitochondrial membrane protein and the acronym MPAN (mitochondrial membrane protein-associated neurodegeneration). In this study we report the clinical description of five patients from five families with NBIA and the subsequent molecular genetic investigation. Five patients diagnosed with NBIA by movement disorders specialists. As required by clinical regulations, all provided written informed consent for genetic testing. The initial diagnosis was based on manifestation of progressive movement disorder and demonstration of a hypointense signal in the globus pallidus on T2-weighted MRI scans. The exons of PANK2, PLA2G6, FAH2 and C19orf12 were amplified by PCR and sequenced. Results. Sequencing of PANK2, PLA2G6 and FAH2 was normal; whereas the molecular analysis of C19orf12 gene revealed a novel homozygous and two heterozygous C19orf12 mutations in patients with NBIA. Further studies are needed to explore the function of C19orf12 in NBIA, and extended genetic analysis of larger patients cohorts will provide more information about the frequency of this disease.

3001F

Mutation of hypoxanthine guanine phosphoribosyltransferase (HPRT), gives rise to Lesch-Nyhan syndrome, which is characterized by hyperuricemia, severe motor disability, and self-injurious behavior, or HPRT-related gout (Kelleys-Geemmiller syndrome). The marked heterogeneity of HPRT deficiency is well known, as reported more than 500 mutations at the HPRT gene (HPRT1) locus. We identified a number of HPRT1 mutations in Asian patients manifesting different clinical phenotypes. In this study, we clarified the mutations of four new Japanese families with the HPRT deficiencies and observed the increase of erythrocyte PRPP concentrations in the deficiencies. A new mutation of G to TT (c.456delinsTT) resulting in a frame shift (p.152fsX154X) in exon 3 has been identified from a Lesch-Nyhan family. In the other Lesch-Nyhan family, a point mutation on intron 7 (c.532+5G>T) causing splicing error (exon 7 excluded) was detected. In two partial deficiency cases with hyperuricemia, two missense mutations of p.D20V (c.59A>T) and p.H60R (c.179A>G) were found, respectively. The mutations of c.532+5G>T and p.D20V were detected firstly in Japanese patients but hitherto has been reported in European families. The p.H60R was reported previ- ously as an asymptomatic missense substitution. The PRPP concentrations in erythrocyte from three Lesch-Nyhan patients (18.5; 33.4; 40.8 nmol/mL) were markedly increased compared with those from normal control (0.8 ± 0.3 nmol/mL). The PRPP concentrations from patients with partial deficiency (1.3 - 7.7 nmol/mL) were also increased, but the degrees were smaller than those of the complete deficiency. The increases of erythrocyte PRPP concentration were correlated approximately with the phenotypic severity. Furthermore, we summarize the spectrum of Asian HPRT muta-

1) Medical genetics, DGM Faculty of biology and medicine, Unil, Lausanne, Switzerland; 2) Service de Neurologie, CHU Henri Mondor, 94000 Créteil, France; 3) CICIs- Health Science Research Centre, Universidade da Beira Interior, 6200-506 Covilhã, Portugal; 4) Solomon H. Snyder Department of Neuroscience, John Hopkins University, Baltimore, MD 21205, USA; 5) Department of Neurodegenerative Disease, Hertie-Institute for Clinical Brain Research and Center for Neurology, 72076 Tübingen, Germany; 6) Institute of Neuropathology, University Hospital RWTH Aachen, 52074 Aachen, Germany; 7) Department of Neurology, University Hospital RWTH Aachen, 52074 Aachen, Germany; 8) Centre de recherche des maladies neuromusculaires, Assistance Publique-Hôpitaux de Paris, CHU Pitié-Salpêtrière, 75013 Paris, France; 9) UnIGENe and Centro de Genética Predictiva e Preventiva, Institute for Molecular and Cellular Biology, 4050 Porto and Serviço de Neurologia, Centro Hospitalar entre Douro e Vouga, 4520-211 Santa Maria da Feira, Portugal; 10) Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (Inserm/UMPC 975; CNRS 7225) and Department of Genetics, Assistance Publique-Hôpitaux de Paris, CHU Pitié-Salpêtrière, 75013 Paris, France; 11) Neurogenetics Group, Ecole Pratique des Hautes Etudes, CHU Pitié-Salpêtrière, 75013 Paris, France; 12) Department of Neurology, Pitié-Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris and University Pierre et Marie Curie, 75013 Paris, France; 13) Department of Neurology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA; 14) Department of Neurology, Friedrich-Baur-Institute, Ludwig-Maximilian-University, 81377 Munich, Germany.

Charcot-Marie-Tooth disease (CMT) comprises a clinically and genetically heterogeneous group of peripheral neuropathies characterized by progressive distal muscle weakness and atrophy, foot deformities, and distal sensory loss. Following identification of two consanguineous families affected by a sive distal muscle weakness and atrophy, foot deformities, and distal sensory neuropathy affecting 1 in 2500 people. The disorder affects both motor and sensory neurons of the peripheral nervous system resulting in the typical abnormalities during Dictyostelium growth and development. CLN3-cells proliferate at an enhanced rate and are able to complete the developmental cycle. However, phenotypes are observed during early, mid-, and late Dictyostelium development suggesting an involvement of CLN3 in cell movement, aggregation, and differentiation. Our current work is focused on expressing human CLN3 in Dictyostelium to see whether its expression can rescue the phenotypes we observe in CLN3-cells. Together, our work has generated a new cellular model for studying CLN3 function in neurodegeneration which will provide new insight into the function of CLN3 in normal and diseased human cells. Our work also further validates the use of Dictyostelium for studying the function of proteins linked to neurodegeneration.

Charcot-Marie Tooth disease (CMT) is the most common inherited peripheral neuropathy affecting 1 in 2500 people. The disorder affects both motor and sensory neurons of the peripheral nervous system resulting in the ‘‘dying back’’ (axonal degeneration) of long nerves. CMT is both clinically and genetically heterogeneous and is traditionally divided into demyelinating (CMT1) and axonal (CMT2) forms based on electrophysiological and neuropathological criteria. For CMT type 2 up to 80% of the genes remain unidentified. CMT research in the last 20 years has made significant discoveries that are elucidating the complexities of motor and sensory nerve biology and the pathogenic process of axonal degeneration. The advance ment of sequencing technologies is providing affordable tools to identify gene mutations in families too small for traditional positional cloning approaches. The strategy of whole exome sequencing (WES) is therefore a very feasible approach for gene discovery in smaller CMT families. We have recently reported a Polish family with CMT type 2 in which all known CMT2 genes have been excluded (1). The family is a three generation kindred with affected individuals in two generations. We have performed WES on four members of the family (3 affected and one normal). Bioinformatic analysis and variant filtering identified a non-synonymous exonic variant in the member of the tubulin gene family. The variant fully segregates with the affected phenotype and has been included in 1400 normal control chromosomes. The variant occurs at a highly conserved amino acid residue of the protein and is predicted to be pathogenic. By querying 164 CMT2 index exomes in the GEMapp database (University of Miami, Miller School of Medicine URL: https://genomics.med.miami.edu/gem.app) we have identified an additional family with a different variant in the same gene. Transient transfection experiments with wild type and mutant expression constructs have shown validation and functional studies supporting pathogenicity of the variant will be presented. This study demonstrates the power of WES for gene discovery in small families, further genetic heterogeneity of CMT type 2 and highlights a pathogenic mechanism altering cellular scaffolding. (1) Kochanski A., Kennerson M. et al. 2005 Neurology 64:533-535.
3005W
Distributions of Degenerative Myelopathy Associated SOD1 Alleles among Privately Owned Dogs. A. Kolicheskii1, R. Zeng1, G.S. Johnson1, M.L. Katz2, T. Mhlanga-Mutangadura1, L. Hansen1, E. Ivensson1, K. Lindblad-Toh3,4, J. Guo5, D.P. O’Brien6, G.C. Johnson1, J.R. Coates2. 1) Department of Veterinary Pathobiology, University of Missouri College of Veterinary Medicine, Columbia, MO; 2) Mason Eye Institute, University of Missouri School of Medicine, Columbia, MO, USA; 3) Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; 4) Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA; 5) Department of Veterinary Medicine and Surgery, University of Missouri College of Veterinary Medicine, Columbia, MO, USA.

Canine degenerative myelopathy (DM) is a common neurodegenerative disease in dogs 8 years-old or older. Similar to human amyotrophic lateral sclerosis patients (ALS) patients, dogs with DM show both upper and lower motor neuron signs at end stage. DM-affected dogs may prove useful as a naturally-occurring disease model to investigate ALS pathogenic mechanisms and to evaluate the efficacy of therapeutic interventions. Two SOD1 mutations, SOD1;c.118G>A and SOD1;c.52A>T, are associated with DM. In this study, we recruited a cohort of SOD1:c.118A/G heterozygous dogs to test the efficacy of therapeutic interventions. Among 3005W privately owned dogs, it should be possible to recruit a cohort of homozygotes to test the efficacy of therapeutic interventions. We recruited a cohort of homozygotes Bernese Mountain Dogs. Four of them developed clinical DM-like signs; whereas, none of the 40 SOD1:c.118A/G heterozygous dogs over 10 years old. Sixteen of 26 SOD1:c.118A homozygotes (62%) exhibited signs of DM when their samples were submitted for genotyping but are currently asymptomatic. We reviewed the clinical history of 10 dogs that were under 8 years-old and without signs of DM when their samples were submitted for genotyping but are currently symptomatic. The SOD1:c.118A allele is found in 33.7% of privately owned dogs including cross-bred dogs and representatives of 37% frequency of the SOD1:c.118A allele in this breed. Twenty-four SOD1:c.118A homozygotes or SOD1:c.118G homozygotes, we reviewed the clinical history of 10 dogs that were under 8 years-old and without signs of DM when their samples were submitted for genotyping but are currently symptomatic. Because the SOD1:c.118A allele is found in 33.7% of privately owned dogs, it should be possible to recruit a cohort of homozygotes to test the efficacy of therapeutic interventions.

3006T

The FIG4 gene encodes a phospholipid phosphatase that removes the 5-phosphate from the signaling lipid PI(3,5)P2, a regulator of late endosome/lysosome vesicle trafficking. Deficiency of PI(3,5)P2 in FIG4 null mice results in enlarged cytoplasmic vacuoles, autophagy defects and neurodegeneration (Ferguson et al, HMG 2012). Hypomorphic mutations of human FIG4 result in the peripheral neuropathy Charcot-Marie-Tooth type 4J (Nicholson et al, Brain 2012). Yunis-Varon syndrome (OMIM 216340) is an autosomal-recessive disorder with neurological and bone involvement including cleidocranial dysplasia and digital anomalies, as well as enlarged vacuoles in neurons, muscle, and cartilage. Using whole-exome and Sanger sequencing, we identified frameshift and missense mutations of FIG4 in affected individuals from three unrelated families with Yunis-Varon Syndrome (Campeau et al, AJHG 2013). The FIG4 genotypes of affected individuals included homozygosity for protein truncation alleles in one family, with heterozygosity for missense and null mutations or homozygosity for missense mutations in the other two families. In a functional assay, the missense substitutions from these families failed to correct the vacuolar defect of Fig4 null mouse fibroblasts, demonstrating loss of function. We identified previously unrecognized skeletal defects in Fig4 null mice, including small skeletons, reduced trabecular bone volume and cortical thickness. Comparison of patients with Yunis-Varon Syndrome and CMT4J indicates that complete absence of functional FIG4 results in central and peripheral nervous system dysfunction, as well as skeletal anomalies, while retention of as little as 0.2% of normal FIG4 activity protects the CNS and causes peripheral nerve impairment only. Neurological findings in null individuals of both species include degeneration with atrophy, atrophy of the cerebral cortex and cerebellar and thalamic nuclei, as well as hypoplasia of the corpus callosum. Our findings demonstrate that null mutations of FIG4 are responsible for a subset of Yunis-Varon Syndrome, the most severe known human disorder caused by defective phosphoinositide metabolism, and demonstrate a previously unrecognized role for PI(3,5)P2 signaling in skeletal development. Studies in progress include genomic analysis of three Yunis-Varon families lacking FIG4 mutations, and further investigation of defective bone development in homozygous null mice.
3007F
SLC25A12 homozygous missense mutation reduces neuronal AGC1 activity and reduced N-acetylaspartate in consanguineous. D. Li1,9, M.J. Falk1,2,3, T. Gollapudi1,2,3, K. Moiseeva1,2,3, J. Gusella1,2,3, T. Fuchs1,2,3, V. Marangoni1,2,3, N. Patrizi1,2,3, P. Lathrop1,2,3, T. Monaco1,2,3, N. Hakonarson1,2,3, E. McCormick2,3, E. Place2,3, F.M. Lasorsa4, F.G. Otieno4, C. Hou4, C.E. Kim5, N. Abdel-Magid5, L. Vazquez5, F.D. Mentic6, R. Chiavacci6, G. Giannuzzi6, E.D. Marsh6, Y. Guo7, L. Tian7, F. Palmieri8,9,10, H. Hakonarson1,2,3,1. The Center for Applied Genomics, Abramson Research Center, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA; 2. Division of Human Genetics Department of Pediatrics, The Children’s Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA; 3. Division of Child Development and Metabolic Disease, Department of Pediatrics, The Children’s Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA; 4. Department of Molecular Pharmacology and Therapeutics, Loyola University Stritch School of Medicine, Mayfield, IL 60153, USA; 5. CNR Institute of Membranes and Bioenergetics, 70125 Bari, Italy; 6. Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, 70125 Bari, Italy; 7. Division of Pediatric Neurology, Departments of Pediatrics and Neurology, The Children’s Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA; 8. Division of Pulmonary Medicine, The Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA; 9. Equal contribution; 10. Corresponding authors.

Whole exome sequencing analysis identified a novel homozygous c.1058G>A (p.Arg353Glu) missense mutation in SLC25A12 that segregated with disease in a consanguineous Indian kindred in which a brother and sister were similarly affected with severely delayed global development, congenital hypotonia, mental retardation, myelination, and notable reduction of the N-acetylaspartate peak on brain spectroscopy. Extensive clinical diagnostic genetic testing was unrevealing, although both affected siblings shared three large homozygous chromosomal regions that were identified by genome-wide microarray analysis. Research-based whole exome sequencing (WES) analysis identified two rare homozygous variants that were predicted to be disease-causing, with only the SLC25A12 mutation found by Sanger validation to segregate with disease in this family. SLC25A12 encodes the neuronal aspartate-glutamate carrier 1 (AGC1), which is a conserved transmembrane protein, which is an essential component of the malate-aspartate shuttle that transfers the reducing equivalents of NADH and H+ from cytosol to mitochondria in neurons. Functional studies demonstrated that the activity of recombinant protein is reduced to fifteen percent of the wild-type rate when measuring either aspartate or glutamate transport. This is the second reported family in which AGC1 has been shown to be causal of a severe, infantile-onset neurologic disease. An AGC1 mutation was identified in the initial case by candidate gene sequencing with fibroblast validation of absent AGC1 activity following identification of severely reduced brain levels of N-acetylaspartate and hypomyelination. Indeed, AGC1 activity enables neuronal export of aspartate, which is the glial substrate necessary for proper neuronal myelination. In summary, these data confirm that autosomal recessive SLC25A12 disease impairs neuronal AGC1 activity and highlight the importance of considering SLC25A12 in the differential diagnosis of infantile-onset seizures, hypotonia, and global developmental delay with abnormal myelination and reduced N-acetylaspartate levels in brain.

3008F
Defect of TARG1/c6orf130 gene causing an autosomal form of familial neurodegeneration with seizure. E. Ozkan1,2, B. Chioza1, M. Patton2, A.G. Ladurner2, G. Timinszky1, I. Ahel3, R. Sharifi3, 1) BMS Genetics, St. George’s University of London, London, United Kingdom; 2) Faculty of Medicine, Bubenandt Institute of Physiological Chemistry, Ludwig Maximilians University of Munich, Munich, Germany; 3) Cancer Research UK, Paterson Institute for Cancer Research, University of Manchester, Manchester, UK.

The genetic causes of neurodegenerative disorders, a heterogeneous group of chronic progressive diseases, remain largely unknown. Here, we report the discovery of TARG1/c6orf130 gene (NM_145063.2) causing severe neurodegeneration with an autosomal recessive trait. The clinical presentations are: progressive global neural dysfunction and severe neurologic development delay, generalized tonic-clonic (grandmal) epilepsy, tetraplegia, absence of tendon reflexes, communication disability, and lack of swallowing reflex. We genotyped the family members on genome-wide SNP microarrays and used the data to determine a single 20 cM homozygosity-by-descent (HBD) locus in region 6p21 (LOD score of 7.4); with performing next generation sequencing we identified the missense change c.227C>T that causes an early stop codon (NP_659500.1:p.R76X) at the conserved C-terminus within TARG1/c6orf130. This gene encodes C6orf130 protein belongs to the macrodomains family which are evolutionarily conserved structural protein involving in adenosine diphosphate (ADP)-ribosylation. Adenosine diphosphate (ADP)-ribosylation is an evolutionarily conserved reversible post-translational protein modification that regulates a wide range of cellular processes, including DNA repair, transcription, telomere dynamics, cell differentiation and proliferation, the inflammatory and immune responses and apoptosis. We identified C6orf130 as a PARP (poly(ADP-ribose) polymerase)-interacting protein that removes mono(ADP-ribosyl)ation on glutamate amino acid residues in PARP-modified proteins. Our biochemical analysis of C6orf130 suggests a mechanism of catalytic reversal involving a transient C6orf130 tyrosyl-(ADP-ribose) intermediate. In human neuron cells, we show that C6orf130 localise to the nucleus. Furthermore, depletion of C6orf130 protein in cells leads to proliferation and DNA repair defects. Collectively, our data suggest that C6orf130 enzymatic activity has a role in the turnover and recycling of protein ADP-ribose polymerase and we have identified the importance of normal neuron cell function in humans. This is for the first time that we show the importance of C6orf130 protein (newly named Terminal ADP-Ribose protein Glycohydrolase; TARG1) in neuron cells by reversing protein ADP-riboaslylation and protein modulation.
Application of high-throughput sequencing to pediatric patients with unresolved diagnoses. R.J. Taft1, C. Simons1, A. Vanderwee2, R.J. Lev- enter4, M.S. van der Knaap3, N.I. Wolf3, R. Schiffmann3, S. Damigni1, P. Pearl1, M. Bloom3, S.M. Grimmmond2, D. Miller2, D.R. Thornbury10, J. Christodoulou10, M. Gabbit11, J. McQuaigh12. 1) Institute for Molecular Bioscience, University of Queensland, St Lucia, Australia; 2) Department of Neurology, Children's National Medical Center, Washington, DC, USA; 3) Department of Neurology, Royal Children's Hospital, Melbourne, Victoria, Australia; 4) Murdoch Children's Research Institute, Melbourne, Victoria, Australia; 5) Department of Child Neurology, VU University Medical Center, Amsterdam, The Netherlands; 6) Institute of Metabolic Disease, Baylor Research Institute, Dallas, TX, USA; 7) Mission Massimo Foundation, Victo- ria, Australia; 8) Department of Pediatrics, Children's National Medical Cen- ter, Washington, DC, USA; 9) Genetic Metabolic Disorders Research Unit, Children's Hospital at Westmead, Sydney, Australia; 10) Murdoch Childrens Research Institute; 11) Genetic Health Queensland at the Royal Brisbane and Women's Hospital, Herston, Queensland, Australia; 12) Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland, Australia.

Rare diseases, the vast majority of which are genetic in origin, have a particularly detrimental effect on children. More than 50% of rare disease patients are paediatric, and at least 30% of these children will not live to see their first birthday. Additionally, many of these patients will be incorrectly diagnosed (at least 40% according to EUROSIDI), or will remain without a final diagnosis. For example, it is well established that half of the patients with leukoencephalopathies, rare central nervous system white-matter disorders, remain a diagnostic mystery despite the fact that they have a genetic aetiology. The obvious need for a large application of high-throughput genome and exome sequencing to paediatric patients with unresolved diagnoses. These include i) identification of the mutations in DARS responsible for an young boy's leukoencephalopathy and the characterisation of a novel disease (HBSL), ii) identification of a de novo TUBB4A mutation responsible for a leukoencephalopathy of unknown genetic aetiology, H-ABC, iii) identification of a novel de novo mutation in KCNT1, a potassium transporter, in a child with non-specific white matter abnormalities and severe epilepsy, which did not respond to specific channel therapies, and iv) less than a month, identification of the mutation responsible for an unexplainable case of Leigh disease, which facilitated patient enrolment in an ongoing clinical trial. In collaboration with the Global Leukoencephalopathies Network, we showed that rarely affected children are now biophenotypically identifiable.

In humans, duplication of Rai1 is associated with the Potocki-Lupski syn- drome, a neurodevelopmental disorder that causes failure to thrive, hypoto- nia, intellectual disabilities and autism. Previously, we had generated a mouse line carrying a transgene encoding Rai1 under the control of a tTa responsive promoter. By crossing this mouse line with camkII-tTa transgenic mice which express tTa downstream of CamkII promoter, we obtained double-transgenic mice (Dt-Rai1) which express transgenic Rai1 specifically in mouse forebrain neurons. Surprisingly, Dt-Rai1 mice showed most of the PTLS-like phenotypes including underweight, lower abdominal fat content, impaired learning and memory ability. Doxycycline admin- istration to Dt-Rai1 transgenic mice (Dt-Rai1) which express transgenic Rai1 specifically in mouse forebrain neurons. Surprisingly, Dt-Rai1 mice showed most of the PTLS-like phenotypes including underweight, lower abdominal fat content, impaired learning and memory ability. Doxycycline admin- istration to Dt-Rai1 transgenic mice.
3013F
The phenotype associated with ASPM mutation expands to severe Seckel syndrome. D.J. Morris-Rosendahl1, G. Vignoli2, E. Mainini2, G. Randazzo2, P. Bonanni2, M. Virdi1, A. Vignoli1, M. Elia2, M.T. Bonati1, L. Larizza1,2,16. 1) Molecular Biology, Institute Auxologico Italiano, Milano, Milano, Italy; 2) Epilepsy and Clinical Neurophysiology Unit, Institute Eugenio Medea, Conegliano Veneto, Italy; 3) Epileptology Center, Hospital F. Benevedelli, Milano, Italy; 4) Epileptology Center, Hospitd San Paolo, Milano, Italy; 5) IRCCS L’Oasi, Troina (EN), Italy; 6) Medical Genetics, San Paolo School of Medicine, University of Milan, Italy.

Seckel syndrome is a very rare (1:15000) neurodevelopmental disease defined by the occurrence of major clinical signs, reported in all patients as psychomotor delay, speech absence, peculiar behavior and ataxia. Frequently other features are present: epilepsy occurring in 80% of AS cases, with a very heterogeneous presentation varying for seizure control and age of onset, and associated to a typical EEG profile, microcephaly and sleep disturbance. Molecular bases of disease consist in genetic and epigenetic defects within 15q11-12 region, which affect the expression of UBE3A gene (MIM 601623) coding for an ubiquitinE3 ligase involved in protein degradation through the ubiquitin-proteasome. Clinical presentation appears heterogeneous, both in the severity of the major features and in the kind and number of less frequent signs. This heterogeneity may only partially be accounted by different genetic defects and by mosaicism occurrence. The most common defect (50-70%) is a deletion originating de novo on the maternally inherited chromosome, with different extension in size. Other mechanisms are 15 paternal uniparental disomy, imprinting defects and point mutations of UBE3A gene. Aiming at disclosing the role of modifier factors, we selected a cohort of 50 AS deleted cases. They were characterized for deletion size and for harboring rare CNVs by microsatellite segregation and SNP array (Human600W-Quad e Omni 1 Quad). In agreement with the literature most cases grouped in the two main deletion classes, BP1-BP3 and BP2-BP3, while a fraction higher than that reported was characterized for deletion size and for harboring rare CNVs by microsatellite segregation and SNP array (Human600W-Quad e Omni 1 Quad). However cases within a specific deletion class displayed a fairly heterogeneous clinical presentation. SNP array results highlighted several variants yet unreported in databases: although no rearrangement was shared among AS patients the collection of variants was highly heterogeneous and different CNVs were disclosed. Most of the affected genes were reported as susceptibility factors for epilepsy, schizophrenia and autism and code for channels implicated in oxidative stress. Validation of the most interesting modifying genes and the resulting genotype-phenotype correlations is ongoing.

3014W
Mutation in the SYNJ1 gene associated with autosomal recessive, early-onset parkinsonism. S. Olgiati1,2, K. Mueller1, G. Borkc3, M. Trimbom3, F. Beleggia1, M. Leipoldt1, G. Nueemberg5, K. Seufert1, B. Wolink2, M. O’Driscoll1. 1) Institute of Human Genetics, Albert-Ludwigs University of Freiburg Medical Center, Freiburg, Germany; 2) Institute of Human Genetics, University of Cologne, Germany; 3) Genom Diabetes and Stetility Centre, University of South Florida, Tampa, Florida, USA; 4) Institute of Human Genetics, University of Ulm, Ulm, Germany; 5) Institute of Medical and Human Genetics, Charité University Berlin, Berlin, Germany; 6) Cologne Genomics Center, Cologne, Germany.

ASPM is one of the 10 genes known to be associated with primary autosomal recessive microcephaly (MCPH), characterized by reduced head circumference at birth and varying degrees of intellectual disability (ID). Mutations in ASPM are responsible for 25-30% of MCPH, depending on the ethnic background of the patients. The phenotypes associated with the MCPH1 genes are continuously being re-defined and mutations in the CENPJ (MCPH6) and CEP152 (MCPH9) genes are known to also cause Seckel syndrome. Clinical findings so far associated with ASPM mutation include late-onset seizures, pyramidal tract involvement, simplified gyral pattern of the cortex, ventricle enlargement, partial corpus callosum agenesis, mild cerebellar hypoplasia, focal cortical dysplasia and unilateral polymicrogyria. Non-neurologic abnormalities include short stature, idiopathic premature puberty and renal dysplasia. We now describe a family with two daughters (P1 and P2) with a severe Seckel syndrome phenotype including severe microcephaly (P1: OFC at birth 25.5 cm, 38th week; P2: OFC at birth 24.5 cm, 37th week), severe brain malformation and short stature (P1 at 8y8m: weight 10 kg, length 1m; P2 at 5y6m: weight 7 kg, length 86 cm). Both girls have severe intellectual disability. P1 has also developed epilepsy, dysmorphism and the absence of premature chromatin condensation, checkpoint responses in EBV-transformed lymphocyte cell lines from P1 and P2A>C in both affected daughters and heterozygosity for a splice mutation, c.688-3G>A in the wild-type allele of the ASPM gene.

3015T
Copy Number Variants and deletion classes as modifier elements of phenotype in Angelman deleted patients. S. Russo1, V. Giorgini1, E. Mainini2, G. Randazzo2, P. Bonanni2, M. Virdi1, A. Vignoli1, M. Elia2, M.T. Bonati1, L. Larizza1,2,16. 1) Molecular Biology, Institute Auxologico Italiano, Milano, Milano, Italy; 2) Epilepsy and Clinical Neurophysiology Unit, Institute Eugenio Medea, Conegliano Veneto, Italy; 3) Epileptology Center, Hospital F. Benevedelli, Milano, Italy; 4) Epileptology Center, Hospitd San Paolo, Milano, Italy; 5) IRCCS L’Oasi, Troina (EN), Italy; 6) Medical Genetics, San Paolo School of Medicine, University of Milan, Italy.

Angelman syndrome is a very rare (1:15000) neurodevelopmental disease defined by the occurrence of major clinical signs, reported in all patients as psychomotor delay, speech absence, peculiar behavior and ataxia. Frequently other features are present: epilepsy occurring in 80% of AS cases, with a very heterogeneous presentation varying for seizure control and age of onset, and associated to a typical EEG profile, microcephaly and sleep disturbance. Molecular bases of disease consist in genetic and epigenetic defects within 15q11-12 region, which affect the expression of UBE3A gene (MIM 601623) coding for an ubiquitinE3 ligase involved in protein degradation through the ubiquitin-proteasome. Clinical presentation appears heterogeneous, both in the severity of the major features and in the kind and number of less frequent signs. This heterogeneity may only partially be accounted by different genetic defects and by mosaicism occurrence. The most common defect (50-70%) is a deletion originating de novo on the maternally inherited chromosome, with different extension in size. Other mechanisms are 15 paternal uniparental disomy, imprinting defects and point mutations of UBE3A gene. Aiming at disclosing the role of modifier factors, we selected a cohort of 50 AS deleted cases. They were characterized for deletion size and for harboring rare CNVs by microsatellite segregation and SNP array (Human600W-Quad e Omni 1 Quad). In agreement with the literature most cases grouped in the two main deletion classes, BP1-BP3 and BP2-BP3, while a fraction higher than that reported was characterized for deletion size and for harboring rare CNVs by microsatellite segregation and SNP array.
3017W A Japanese girl with severe form of vanishing white matter disease resembling Cree leukoencephalopathy. K. Takano1, T. Wada1, H. Osaka2, Y. Tsurusako2, H. Saito3, N. Matsumoto1. 1Division of Neurology, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2Human Genetics, Yokohama City University, Yokohama, Kanagawa, Japan. Vanishing white matter disease (VWM)/childhood ataxia with central hypomyelination (CACH) is an autosomal recessive brain disorder caused by mutations in each of the 5 genes (EIF2B1-5), encoding the 5 subunits of eukaryotic translation initiation factor 2B (eIF2B). Brain magnetic resonance imaging (MRI) shows progressive rarefaction and cystic degeneration of the white matter and cerebellar ataxia. The clinical phenotype is characterized by early childhood onset and chronic progressive neurological deterioration with cerebellar ataxia, spasticity, optic atrophy and epilepsy. Recently, it has become apparent that the onset of disease varies from antenatal period to adulthood. The current study aimed to report our experience with a case of Cree leukoencephalopathy and describe de novo. We performed a prenatal diagnosis in a candidate family by fluorescence in situ hybridization, 421 cases of suspected inherited peripheral neuropathies were enrolled from Japan. Mutation screening was initially performed on a custom DNA microarray (Affymetrix) containing all 28 known CMT disease-related genes. The negative cases subsequently proceeded to exome sequencing with Illumina HiSeq2000. Using Sanger's method, we confirmed the suspected variants, and a segregation study was performed if the variants were novel. [Result] In 421 cases of suspected CMT, causative mutations were discovered in 56 cases (13.3%) using DNA microarray. Exome sequencing then revealed another 28 reported mutations in the known CMT disease-related genes. Correspondingly, the positive rate of mutation detection was increased in denaturing CMT (from 15.0% to 23.9%), and axonal CMT (from 12.2% to 22.4%) as well. The total positive rate of mutation detection was improved from 13% to 20%. [Conclusion] We compared the ability of DNA microarray analysis and exome sequencing to detect known pathogenic mutations in CMT, and exome sequencing proved to be more effective. Next-generation sequencing is a trustworthy method in high-throughput genetic analysis. However, no pathogenic mutation was found in approximately 80% of cases with suspected CMT. In future, we can expect that a large number of novel disease-related genes will be discovered.

3018T Proteolipid protein 1 and gap junction α12 gene mutations in 72 Chinese patients with Pelizaeus-Merzbacher disease/ Pelizaeus-Merzbacher like disease like disease and prenatal diagnosis of 15 fetuses in twelve Chinese families with PMD probands. J. Wang1, D. Li2, Y. Wu3, J. Xiao4, Q. Gu5, H. Zhao6, J. Jiang2, Y. Yang7, X. Bao8, H. Xiong9, Y. Zhang10, T. Ji11, M. Li12, Y. Meng2, F. Fang3, Z. Niu13, J. Qin14, H. Shi15, F. Zhang11, X. Wu16, Y. Jiang17. 1) Peking University First Hospital, Beijing, China; 2) Department of Image, Peking University First Hospital, Beijing 100034, China; 3) Department of Neurology, Renji Hospital affiliated to Shanghai Jiao Tong University, Shanghai 200010, China; 4) Department of Neurology, Shanyi Dayi Hospital Affiliated to Shansi Medical University, Taiyuan 030001, China; 5) Department of Pediatrics, Peking Union Medical College Hospital, Beijing 100730, China; 6) Department of Neurology, Beijing Children's Hospital Affiliated to Capital Medical University, Beijing 100045, China; 7) State Key Laboratory, School of Life Science, Fudan University, Shanghai 200433, China.

Purpose The object of this study was to identify proteolipid protein 1 (PLP1) and gap junction α12 mutations in 72 Chinese patients (P1-72) with Pelizaeus-Merzbacher disease (PMD)/PMLD and prenatal diagnosis of fifteen fetuses in twelve Chinese families with PMD probands. Methods Genomic DNA was extracted from peripheral blood samples. At 19 or 20 weeks gestation, amniotic fluid/chorionic villus sampling (AFS/CVS) was performed. Gene dosage was determined by Multiplex Ligation-dependent Probe Amplification (MLPA). All 7 exons and exon-intron boundaries of PLP1 gene were sequenced by Sanger’s method. Results Of these 72 patients, there were 18 transitional, 45 classical, and 9 congenital PMD according to the clinical and radiological presentation. PLP1 duplications were identified in patients 1-52 with PMD, accounting for 72.2% (52/72). Their mothers were PLP1 duplications carriers. 15 hemizygous missense mutations including eight novel mutations and one reported splicing mutation (IVS5-1G>A) were found in 17 Patients (P33-69) with PMD (23.6%), 12 out of 17 for their mother were showed to be a heterozygote of those mutation, and the remains 5 patients demonstrated de novo mutations. For all patients with PMD, we performed in situ hybridization and real-time PCR using probes of exon 1 and 2 of PLP1 for 24 cases with PLP1 duplication mutations using fluorescence in situ hybridization, 421 cases of suspected inherited peripheral neuropathies were enrolled from Japan. Mutation screening was initially performed on a custom DNA microarray (Affymetrix) containing all 28 known CMT disease-related genes. The negative cases subsequently proceeded to exome sequencing with Illumina HiSeq2000. Using Sanger’s method, we confirmed the suspected variants, and a segregation study was performed if the variants were novel. [Result] In 421 cases of suspected CMT, causative mutations were discovered in 56 cases (13.3%) using DNA microarray. Exome sequencing then revealed another 28 reported mutations in the known CMT disease-related genes. Correspondingly, the positive rate of mutation detection was increased in denaturing CMT (from 15.0% to 23.9%), and axonal CMT (from 12.2% to 22.4%) as well. The total positive rate of mutation detection was improved from 13% to 20%. [Conclusion] We compared the ability of DNA microarray analysis and exome sequencing to detect known pathogenic mutations in CMT, and exome sequencing proved to be more effective. Next-generation sequencing is a trustworthy method in high-throughput genetic analysis. However, no pathogenic mutation was found in approximately 80% of cases with suspected CMT. In future, we can expect that a large number of novel disease-related genes will be discovered.

3019F Comprehensive Mutation Analysis of 421 Cases with Charcot-Marie-Tooth Disease using Microarray and Exome Sequencing. A. Yoshiura1, A. Hashiguchi2, Y. Higuchi3, J. Yuan4, Y. Okamoto5, T. Nakamura6, J. Mitsu7, H. Ishiura8, Y. Takahashi9, J. Yoshimura10, K. Doi11, S. Morishita12, S. Tsuji13, H. Takashima14. 1) Kagoshima University, 8-35-1 Sakuragaoka, Kagoshima City, Japan; 2) University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan.

[Objective] Charcot-Marie-Tooth (CMT) disease comprises a group of clinically and genetically heterogeneous inherited peripheral neuropathies. In our study, to identify causative genes and new molecular mechanisms of CMT, DNA microarray and exome sequencing were successively applied. [Methods] From April 2005 to December 2011, after excluding the PMP22 duplication mutation using fluorescence in situ hybridization, 421 cases with suspected inherited peripheral neuropathies were enrolled from Japan. Mutation screening was initially performed on a custom DNA microarray (Affymetrix) containing all 28 known CMT disease-related genes. The negative cases subsequently proceeded to exome sequencing with Illumina HiSeq2000. Using Sanger’s method, we confirmed the suspected variants, and a segregation study was performed if the variants were novel. [Result] In 421 cases of suspected CMT, causative mutations were discovered in 56 cases (13.3%) using DNA microarray. Exome sequencing then revealed another 28 reported mutations in the known CMT disease-related genes. Correspondingly, the positive rate of mutation detection was increased in denaturing CMT (from 15.0% to 23.9%), and axonal CMT (from 12.2% to 22.4%) as well. The total positive rate of mutation detection was improved from 13% to 20%. [Conclusion] We compared the ability of DNA microarray analysis and exome sequencing to detect known pathogenic mutations in CMT, and exome sequencing proved to be more effective. Next-generation sequencing is a trustworthy method in high-throughput genetic analysis. However, no pathogenic mutation was found in approximately 80% of cases with suspected CMT. In future, we can expect that a large number of novel disease-related genes will be discovered.

3020W ABNORMAL COPY NUMBER VARIANTS ARE FREQUENT IN PATIENTS WITH MALFORMATIONS OF CORTICAL DEVELOPMENT ASSOCIATED WITH EPILEPSY. F. Torres1, D.A. Souza2, M.M. Guerreiro2, M.A. Monte negro3, A.C. dos Santos2, V.C. Terra2, A.C. Sakamoto2, F. Cendes2, I. Lopes-Cendes2. 1) Department of Medical Genetics, UNICAMP, Campinas, Sao Paulo, Brazil; 2) Department of Neurology, UNICAMP, Campinas, Sao Paulo, Brazil; 3) Department of Neurosciences, USP, Ribeirao Preto, Sao Paulo, Brazil.

Patients with malformations of cortical development (MCD) often suffer from seizures which are frequently refractory to treatment with antiepileptic drugs. Advances in molecular genetics have led to a better understanding of the mechanisms underlying several types of MCDs. However, mutation events remain unidentified in the majority of patients. Recent studies have implicated large, rare copy number variants(CNVs) in a range of neurodevelopmental disorders. Therefore, the aim of this study was to investigate whether CNVs could be involved in different types of MCDs. We used a high resolution SNP-array, CytoScan® HD, to investigate CNVs in a cohort of 33 patients with MCDs, including lissencephaly spectrum, periventricular nodular heterotopia, Schizencephaly, and neuronal migration, have been previously reported in patients with 

ARX

(3020W) and neuronal migration, have been previously reported in patients with 

DAAM1

(3020W) and neuronal migration, have been previously reported in patients with 

HAUS7

(3020W) and neuronal migration, have been previously reported in patients with 

EIF2B1-5

(3020W) and neuronal migration, have been previously reported in patients with 

ABNORMALCOPYNUMBERVARIANTSAREFREQUENTINPATIENTS

with MCDs. However, no pathogenic mutation was found in approximately 80% of cases with suspected CMT. In future, we can expect that a large number of novel disease-related genes will be discovered.
**3021T**


1) Département de Biochimie et Génétique, CHU Angers, CNRS UMR6214, INSERM UMR1083 France; 2) Département d’Ophthalmo-logie CHU Angers, France; 3) Département d’Ophthalmo-logie CHU Lille, France; 4) Département d’Oto-Rhino-Laryngologie, CHU Lille, France; 5) Département de Biochimie et Génétique, CHU Rouen, France; 6) Service de Génétielle Médicale, Centre Hospitalier Arras, France; 7) INSERM U1051, Montpellier, France; 8) Département d’Oto-Rhino-Laryngologie, CHU Montpellier; 9) National Eye Clinic, Kennedy Center and University of Copenhagen, Denmark; 10) Département de Neurologie, CHU Angers, France.

Heterozygous mutations of OPA1 (Optic Atrophy 1), that encodes for a dynamin-related GTPase involved in mitochondrial DNA maintenance, are linked to autosomal dominant optic atrophy (DOA). Optic neuropathy usually occurs insidiously in the first decade of life, about 20% of patients later developing extra-ocular symptoms (DOA+) in particular deafness. We retrospectively reviewed the files of 1380 DOA patients referred to our laboratory from 2003 to 2011. OPA1 mutation was identified in 327 patients (24%) of whom 21 patients (6.4%) had hearing impairment. In 10 patients deafness was detected under age 20 (~48%), in 3 patients over age 20 (~14%) and in 8 patients the age of onset was unknown (38%). Molecular screening of OPA1 identified three mutations, p.Arg445His, p.Gly401Asp and p.Leu243 previously reported associated to optic atrophy and hearing loss; two mutations, p.Val291_Phe328del and p.Ile463_Phe464dup reported in isolated deafness are the only signs in 13/21 patients (62%) additional neurological signs are present in 8/21 patients (38%). In 54% of patients hearing loss started prior to visual abnormalities. A particularly careful should be performed in patients carrying an OPA1 missense mutation because there is a 2-4 fold increased risk of developing DOA plus with missense mutations affecting the GTPase domain or the dynamin domain. In conclusion, deafness linked to OPA1 is probably underestimated since audiological investigations are performed in only a minority of OPA1 mutation carriers.

**3022F**


1) INSERM U781, Université Paris Descartes, Sorbonne Paris City, Institut IMAGINE, PARIS, France; 2) Université Paris-Sud, CNRS, UMR 8621, Institut de Génétique et Microbiologie, Orsay, F-91405, France; 3) Plateforme de Génomique, Institut IMAGINE, 75015 Paris, France; 4) Plateforme de Bioinformatique, Université Paris Descartes, Sorbonne Paris Cité, Institut IMAGINE, 75015 Paris, France; 5) Unit on Human Copper Metabolism, Molecular Medicine Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, USA; 6) Department of Paediatric Radiology, Hôpital Necker Enfants Malades, APHP, Paris, France.

We investigated a large pedigree consistent with X-linked inheritance by a combination of linkage analysis and X-exome sequencing using RainDance Technologies®. All 6 affected males exhibited epileptic encephalopathy. Epilepsy onset ranged from 2 to 22 months and consisted of myoclonic seizures, febrile seizures and Lennox-Gastaut syndrome. One variant perfectly segregating with the disease was identified. This variant (c.3670G>A; p.Gly1224Ser) is located in the gene encoding the P-type copper-transporter ATPase (ATP7A). Yeast complementation assays using the S. cerevisiae copper transport mutant ccc2 demonstrated its pathogenicity since the p.Gly1224Ser mutant ATP7A cDNA was unable to restore normal growth on nutritionally restricted media. Mutations in ATP7A yield three distinct X-linked conditions. Menkes disease (MD) presents in the first months of life with failure to thrive, developmental delay, and seizures. The diagnosis of MD is usually confirmed by decreased serum copper and ceruloplasmin levels and intracellular copper retention. Occipital horn syndrome features more subtle developmental delay, dysautonomia, and connective tissue anomalies beginning in early childhood. Finally, ATP7A-related distal motor neuropathy is characterized by a late-onset neurological phenotype that resembles Charcot-Marie-Tooth disease, type 2.

Because none of these diagnoses had been suggested in our family we undertook detailed re-evaluation of the patients. Brain MRI showed normal cerebral blood vessels morphology but metal overload in basal ganglia. Despite normal hair texture, optic microscopy revealed rare pili torti. Biochemical analysis showed not only normal ceruloplasmin and mildly increased serum copper levels but surprisingly, low intracellular copper level in patient’s fibroblasts compared to controls (15.5 ng/mg of protein). Finally, immunohistochemical analyses performed in patient’s fibroblasts showed normal localization to the trans-Golgi network (TGN) of the mutant protein and normal trafficking from the TGN to the plasma membrane in response to increased copper concentration. Our report further expands the phenotypic spectrum associated with ATP7A mutations. Additional experiments are now underway to elucidate the pathophysiological mechanism underlying this condition which is likely different from known MD/OHS-causing mutations.
3023W Phenotypic and transcriptomic characterization of the RPE affected by mutations that cause RNA splicing factor retinitis pigmentosa, M.H. Farkas1, E.F. Nandrot2, E.D. Au, D. Lew2, K. Bujakowska1, M.E. Souza1, D.G. Taub1, S.S. Bhattacharya2, E.A. Pierce2. 1) Occlar Genomics Institute, Berman-Gund Laboratory, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA; 2) Institut de la Vision, Paris, France.

Mutations in the Pre-mRNA Processing Factors 3, 8, and 31 (PRPF3, 8, and 31) cause non-syndromic retinitis pigmentosa (RP) in humans, an inherited retinal dystrophy (IRD). It is currently unclear what mechanisms, or which tissues and when mutations are present if mutations are expressed proteins. Mice with the human mutations in these genes show limited vision loss, accompanied by late-onset morphological changes in the retinal pigment epithelium (RPE), specifically a loss of basal infoldings, vacuole accumulation, and DNA expression. Thus, RPE is critically important for the overall health and maintenance of the retina, we set out to determine if the observed morphological changes are preceded by abnormal function. Specifically, we investigate, in detail, the phagocytic mechanism of the RPE in Prp3, Prp8, and Prp31-mutant mice. Phagocytosis deficiencies of up to 50% are observed in primary RPE cultures, along with a nearly 60% decrease in RPE adhesion. The RPE diurnal rhythm is also affected in all 3 models. In addition, RNA-Seq data show alterations in gene expression in RPE and retina, as well as in brain and muscle. The greatest number of transcripts found to be differentially expressed (DE) are associated with the RPE of the Prp3-mutant mice, with over 2000 identified. In comparison, all other tissues/models had a few hundred DE transcripts. Categorization of DE transcripts in the RPE of all 3 models revealed an over abundance of transcripts that affected RNA splicing factors and RNA processing genes. Interestingly, there is no overlap of DE transcripts among all 3 models. All 3 mutant models also display an increase in unique novel splicing events, and we find differential expression of novel genes between the mutant and control models. Our findings appear to be the only pathway to the RNA splicing factor forms of RP. While degenerative changes appear in 2-year-old mice, consistent with the late disease onset observed in humans, our characterization suggests the RPE is dysfunctional soon after birth. RNA-Seq analyses, to date, suggest that aberrant splicing occurs globally, yet only affects retinal tissue. The divergent altered transcripts among the 3 models suggest different mechanisms of pathogenesis, even though all 3 Prp proteins are found in the same core component of the spliceosome. These results provide the greatest insight into the pathogenesis and mechanism of RNA splicing factor RP to date.

3024T THE RECRUITMENT OF RNA POLYMERS AS IS MPREDICTED IN CORNEAL DISEASES. A. Musio1, S. Blodseau2, C. Amato2, V. Quarantotti1, F. Cucco1,2, J. Krantz2, L. Mannini1. 1) Istituto di Ricerca Genetica e Biomedica, CNR, Pisa, Italy; 2) Centre de recherche en cancérologie de l’Université Laval and Centre de recherche du CHU de Québec (Hôpital-Dieu de Québec), Département de biologie moléculaire, Faculté de Médecine, Université Laval, Québec, Canada; 3) Dipartimento di Biologia, Università di Pisa; 4) Division of Human Genetics, The Children’s Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, US.

Cohesin ensures correct chromosome segregation by holding sister chromatids together from the S phase until their separation in anaphase. Mutations in cohesin and regulatory cohesin genes, NIPBL, SMCA1A, HDAC8 and SMCC1A, cause autosomal recessive and X-linked achondroplasia and short stature syndromes, respectively. Among them S549N, L69H in classical CF and G126S, Y852F in CBDAV/CUAVD males were among the rare missense mutations (Sharma et al., 2009). In this study we have attempted to conduct in vitro gene expression analysis to establish genotype and phenotype correlation and to characterize these four rare missense mutations according to the mechanism that disrupts cohesin function protein. All four mutations from Indian population were characterized by expressing pEGFP-CFTR constructs in BHK-21 cells via 3 step technique viz; CFTR cellular localization was determined by confocal microscopy, where as Western blot analysis and automated iodide efflux assays was used to determine CFTR maturation processes and its channel activity respectively. In Western blot analysis only b-band is obtained for L69H substitution as for F508del whereas in the case of other mutants both b and c bands were found, indicating L69H mutation impair CFTR maturation process, the finding was again confirmed by confocal imaging, iodide Efflux assay reveled significant decrease in channel activity for L69H and S549N mutants CFTR expressing cells in comparison to WT, although this decrease in channel activity was rescued when cells were incubated at 27 °C. When the effect of CFTR correctors was checked, it was found that VX809 significantly ameliorate the defect caused by L69H mutation. Y852F and G126S have no impact on CFTR maturation and function. In conclusion L69H mutation is a class II CF mutation causing impaired maturation of CFTR and decreases the probability of function like F508del mutation. This defect is rescued by the corrector VX 809. Whereas S549N mutation can be categorized into a class II/III mutation causing impaired maturation and reduced channel activity.

3025F Characterization of Novel mutations found in Classical and Infertile CF males of Indian population: A molecular approach to establish genotype and phenotype correlation. P. Prasad1, N. Dhawan2, S. K. Sharma1, F. Becq2. 1) Department of Biochemistry, PGIMER, Chandigarh, India; 2) Institute of Physiology and Cell Biology, University of Poitiers, France.

Cystic Fibrosis an autosomal recessive disorder is usually considered as most common disease for Indian population hence much less is known about this disease in context with Indian sub continent. Recently we have established a spectrum of mutations in CFTR gene from classical Cystic Fibrosis as well as from infertile male patients with CBDAV/CUAVD in Indian populations. Among them S549N, L69H in classical CF and G126S, Y852F in CBDAV males were among the rare missense mutations (Sharma et al., 2009). In this study we have attempted to conduct in vitro gene expression analysis to establish genotype and phenotype correlation and to characterize these four rare missense mutations according to the mechanism that disrupts CFTR protein function. All four mutations from Indian population were characterized by expressing pEGFP-CFTR constructs in BHK-21 cells via 3 step technique viz; CFTR cellular localization was determined by confocal microscopy, where as Western blot analysis and automated iodide efflux assays was used to determine CFTR maturation processes and its channel activity respectively. In Western blot analysis only b-band is obtained for L69H substitution as for F508del whereas in the case of other mutants both b and c bands were found, indicating L69H mutation impair CFTR maturation process, the finding was again confirmed by confocal imaging, iodide Efflux assay reveled significant decrease in channel activity for L69H and S549N mutants CFTR expressing cells in comparison to WT, although this decrease in channel activity was rescued when cells were incubated at 27 °C. When the effect of CFTR correctors was checked, it was found that VX809 significantly ameliorate the defect caused by L69H mutation. Y852F and G126S have no impact on CFTR maturation and function. In conclusion L69H mutation is a class II CF mutation causing impaired maturation of CFTR and decreases the probability of function like F508del mutation. This defect is rescued by the corrector VX 809. Whereas S549N mutation can be categorized into a class II/III mutation causing impaired maturation and reduced channel activity.

3026W Distinct prevalence of homozygous p.V37I variant of GJB2 in Chinese Hans with severe-to-profound, mild-to-moderate or normal hearing phenotype. Y. Chai1,2, Y. Tao1,3, H. Wu4,5. 1) Ear Institute, Shanghai Jiao-Tong University School of medicine, Shanghai, China; 2) Department of Otolaryngology-Head and Neck Surgery, Xinhua Hospital, Shanghai Jiao-Tong University School of medicine, Shanghai, China.

The p.V37I variant of GJB2 is highly prevalent in East Asians, with a carrier frequency of 6.2% in Chinese Hans. It has been shown that homozygous p.V37I may lead to mild-to-moderate hearing impairment and is a genetic risk for postnatal childhood hearing impairment. On the other side, the complete spectrum of hearing phenotypes associated with the p.V37I variant remains elusive. Recently, in 80 probands with severe-to-profound and mild-to-moderate childhood hearing impairment, respectively, as well as 1000 Chinese Han adults with normal hearing. Direct sequencing of all subjects for GJB2 showed that a significant percentage (12.5%, 11/88) of probands with mild-to-moderate hearing impairment was homozygous for p.V37I, in contrast to 0.1% (1/1000) in adults with normal hearing (P<0.001). Homozygous p.V37I was also moderately over-represented in probands with severe-to-profound hearing impairment (16.3%, 14/88, P<0.01). Targeted next-generation sequencing of 79 known deafness genes in those 11 probands, however, identified one proband with homozygous p.Y199SX mutation in CDH23 in addition to the homozygous p.V37I variant in GJB2. Our results suggested that the highly penetrant p.V37I variant is associated with a broad spectrum of hearing phenotypes. Though present most frequently in patients with milder degree of hearing impairment, it may also exhibit much severer or completely normal hearing phenotypes. Cautions should be taken when interpret the pathogenic causes for patients carrying the homozygous p.V37I variant though, as this variant may have separate pathogenic causes in addition to a non-penetrant homozygous p.V37I.
Mosaic Missense Mutations in the RNase IIIb Domain of DICER1 Cause JANUS, a Novel Overgrowth Syndrome. S. Klein1,2, H. Lee1,2, S. Grigelioniene1, D. Nilsson1, M. Pettersson1, B. M. Andersen1,2, P. Gustavsson1, H. Valta2, S. Geiberger3, M. Nordenskjöld4, M. Pettersson5, K. Lagerstedt5, F. Taylair1, J. Wincent1, T. Laurell1, M. Pekkinen5, M. Nordenskjöld1, O. Mäkitie1,6, A. Nordgren1, 1) Department of Molecular Medicine and Surgery and Center of Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Department of Women’s and Children's Health Karolinska Institutet, Stockholm, Sweden; 3) Department of Pediatric Radiology, Karolinska University Hospital Solna, Stockholm, Sweden; 4) Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden; 5) Children’s Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; 6) Folkhälans Institute of Genetics, Helsinki, Finland.

Point mutations in PDE4D have recently been linked to acrodysostosis, an autosomal dominant disorder characterized by facial dysostosis, severe brachydactyly, nasal hypoplasia, short stature, intellectual disability and hormone resistance. We found PDE4D mutations in five unrelated patients with acrodysostosis. In addition, two patients with large heterozygous deletions encompassing the entire PDE4D locus were identified. These patients do not have acrodysostosis but suffer from a novel intellectual disability syndrome with severe anorexia and characteristic facial features including a long nose and a small chin. Careful clinical comparisons between these patients raised the possibility that different types of PDE4D mutations result in opposite phenotype in humans. To further explore the function and pathogenicity of the identified alleles we used zebrafish embryos. We found that zebrafish pde4d was expressed in the brain of the developing embryo. Embryos injected with two different morpholinos showed a consistent phenotype (short body length, large head, small jaw), as did embryos injected with mRNAs harboring acrodysostosis point mutations (curved body, small head, large jaw). Interestingly, the jaw of embryos injected with mRNA was severely enlarged while in embryos where pde4d was suppressed the jaw size decreased, hence capturing the mirror phenotype seen in humans. In addition, results from co-injections of mutated mRNA and wild-type mRNA showed a more severe phenotype compared to embryos injected with the mutated mRNA alone when the mutations were located in the upstream conserved regions of the PDE4D protein, while this effect could not be observed when the mutations were located further down, in the catalytic domain of the protein. In summary, haploinsufficiency of the PDE4D gene results in a previously undescribed syndrome with several characteristic traits that represent the opposite of what we see in patients with acrodysostosis. By comparing overexpression of human mutated transcripts to pde4d knockdown in zebrafish embryos we could successfully replicate this mirror phenotype. Our data also suggest that the genetic mechanism of the dominant mutations causing acrodysostosis might be different depending on where in the protein they are situated.
3029W
Loss of function mutations in Carbonic Anhydrase XII result in hyponatraemic dehydration and elevated sweat chloride concentration. S. Vecchio-Pagán1, M. Lee1, N. Sharma2, A. Waheed2, D. Belchis2, J. Hertecant3, W. Sty2, G.R. Cutting1, 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Medicine, Baltimore, MD; 2) Department of Biochemistry, St. Louis University School of Medicine, St. Louis, MD; 3) Pathology, Johns Hopkins School of Medicine, Baltimore, MD; 4) Tawam Hospital, UAE National University, Al Ain United Arab Emirates.

Functional analysis can provide key evidence that a gene linked to a rare disease by exome sequencing is causative; however, the appropriate physiological context needs to be considered when testing the functional effect of mutations. We provide an illustrative example that demonstrates a new role for CA XII mediated bicarbonate metabolism in determining the ionic composition of sweat. Exome sequencing revealed a novel homozygous mutation (H121Q) in carbonic anhydrase XII (CA XII) in 2 siblings exhibiting hyponatraemic dehydration, salt craving, and elevated sweat chloride (Cl−) levels in a consanguineous Omani pedigree with three normal siblings. CA XII maintains proper physiological pH through the reversible dehydrogenation of HCO3− to CO2 and H2O at its extracellular catalytic site. To assess the functional effect of H121Q, we measured the rate of CO2 hydration. In this assay, H121D displayed reduced activity (15.4±3.6 % WT activity, n=11). Prior publications have described a different point mutation in CA XII, E143K, in an Israeli Bedouin tribe which segregated with an autosomal recessive phenotype similar to that of the Omani family (OMIM#143860). However, the reported functional assays revealed that the E143K mutation only reduced CA XII function to ~70% WT activity, which we confirmed (75.5±4.9 %WT, n=11). To reconcile the apparent dramatic difference in function between these two mutations, we sought to determine whether either led to altered cellular localization of CA XII. Confocal microscopy using a Protein Tech anti-CA XII antibody was performed upon polarized MDCK cells expressing WT and both CA XII mutants. Wild-type, H121Q and E143K CA XII showed similar localization, primarily at the basolateral cell membrane, suggesting that the catalytic sites of these mutants retain an extracellular orientation. As the E143K mutant had previously been reported to have altered anion sensitivity, we assayed both mutants under Cl− concentrations with normal physiological intracellular pH. Cells expressing CA XII in 2 siblings autosomal recessive pedigrees produce severe loss-of-function in CA XII activity.

3030T

Incontinentia Pigmenti (IP, OMIM308300) is a X-linked dominant neurocutaneous disease associated with skin defects and with extracutaneous manifestations at variable frequency. IP is caused by mutations in NEMO encoding the regulatory subunit of the IKK complex responsible of generating missense mutations. Then, we produced key discoveries on the altered nucleotide sequences result in a premature stop-codon and 10% altered nucleotide sequences in IP patients.

NEMO mutations by exploring with high accuracy both genomic context in the disease area, and intronic sequences. Of the proteases which show overactivity in cutaneous inflammation and allergy including high expression of IL-1α, TNF-α and the pro-Th2 cytokine TSLP with mast cells and eosinophils infiltration. Overtime, persistent scratching leads to alopecic, erosive and crusty skin lesions with lymph nodes hyperplasia. Inflammation was also detected at the systemic level, with high serum IgE and TSLP levels. CD4+ T cells from lymph nodes showed increased secretion of Th2 and Th17 cytokines. These results help deciphering the role of KLK5 in the NS phenotype and provide a useful model for the development of new therapeutic strategies.

3031F
Skin-specific Kallikrein-5 transgenic mice recapitulate the main features of Netherton syndrome and provide a viable model for therapeutic approaches. L. Furio1,2, S. de Veer1, T.C. Deraison1, A. Brott1,2, C. Bonnart1, A. Robin1,2, A. Hovnanian1,2, 1) University Paris Descartes Sorbonne Cite, Paris, France; 2) INSERM U781 and Imagine Institute of Genetic Diseases, Paris, France; 3) Department of Genetics, Necker hospital, Paris, France.

Netherton syndrome (NS) is a severe genetic skin disease caused by mutations in SPINK5 encoding the lymphoepithelial Kazal-type-related serine protease inhibitor (LEKTI). Of the proteases which show overactivity in NS (kallikrein-related peptidase (KLK) 5, KLK7 and elastase 2), KLK5 is a potential key initiator of the proteolytic cascade. To address the role of KLK5 in the disease, we have generated a transgenic murine model overexpressing human KLK5 in the skin (TgKLK5). The transgene was highly expressed in the granular layer of TgKLK5 epidermis and was associated with increased proteolytic activity. In situ zymography showed enhanced KLK5 activity, as well as elevated KLK7 and elastase 2 activities, two proteases which have been proposed to be activated by KLK5. TgKLK5 mice show low weight at birth with a growth delay. They show a skin barrier defect with increased transdermal water loss. The stratum corneum is detached from the underlying epidermis as a result of desmosomal cleavage. Whiskers and hairs are abnormal and reduced. Additionally, TgKLK5 mice displayed hallmarks of cutaneous inflammation and allergy including high expression of IL-1α, TNF-α and the pro-Th2 cytokine TSLP with mast cells and eosinophils infiltration. Overtime, persistent scratching leads to alopecic, erosive and crusty skin lesions with lymph nodes hyperplasia. Inflammation was also detected at the systemic level, with high serum IgE and TSLP levels. CD4+ T cells from lymph nodes showed increased secretion of Th2 and Th17 cytokines. These results help deciphering the role of KLK5 in the NS phenotype and provide a useful model for the development of new therapeutic strategies.
3033T
Clinical whole exome sequencing identifies mutations in UPK3A in individuals with renal adysplasia. M.R. Bekheimia1,2, Z. Niu1, D.A. Scott1, L. Polocko1,3, P. Lupo1, J.W. Belmont1, D.J. Lamb2,3, Y. Yang1, C.M. Eng1. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Center for Reproductive Medicine, Baylor College of Medicine, Houston, TX; 3) Scott Department of Urology, Baylor College of Medicine, Houston, TX.

Congenital Anomalies of Kidney and Urinary Tract (CAKUT) are the most important causes of pediatrics renal failure. While missense mutations in UPK3A gene are rare cause of renal adysplasia, nonsense mutations were not previously reported to be associated with the renal phenotype. Clinical whole exome sequencing is performed in the Medical Genetics Laboratory at Baylor College of Medicine as a clinical test. This study reports our observation of UPK3A mutations during the clinical sign-out of the requested cases. ACMG standards are used for classification of the genetic variants. We have identified a heterozygous (c.545G>A, p.W182X) nonsense mutations in UPK3A in 3 patients with a clinical diagnosis of either unilateral renal agenesis (2 patients) or renal dysplasia (1 patient). Both patients with unilateral renal agenesis inherited the p.W182X variant from an apparently unaffected mother. A parental sample was not available for analysis of the third patient who had renal dysplasia. All 3 patients had other organ involvement. We also identified the p.W182X variant in 4 more individuals who apparently did not have any kidney phenotype. Twelve missense, 2 splice site (c.571+1G>A) and one indel variants were also detected. Among the 12 patients with missense variants, one had echogenic kidneys together with prematurity, dysmorphic features and developmental delay. Of the 2 patients with a splice site variant (c.571+1G>A), 1 had echogenic kidneys and gonadal agenesis on pelvic ultrasound. While there were 2 earlier reports of missense mutations in UPK3A found in association with renal adysplasia, this is the first report of other types of mutations, such as nonsense variants, that are likely to be related to the renal adysplasia phenotype. Thus, a possible loss of function mechanism resulting from a mutation in UPK3A may cause renal adysplasia. Although the nonsense variants display a higher penetrance than the missense variants, in 2/3 of affected cases the variants were inherited from apparently unaffected parents. The role of the uropodin gene mutations in Cakut phenotype clearly requires further investigation. Acknowledgement: Supported in part by K12 DK0083014, the Multidisciplinary K12 Urologic Research (KURe) Career Development Program to DJL (MRB is a KURe Scholar).

3035W
Variable phenotype in individuals with a MUC1 mutation causing medullary cystic kidney disease type1. S. Knoch1, A.J. Bleyer1, K. Kidd2, K. Hodanova1, M. Zuna1, P. Vyletal1, H. Hartmannova1, V. Stranecky1. 1) Inst Inherited Metabolic Dis, Charles Univ 1st Faculty, Prague 2, Czech Republic; 2) Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC 27157.

A cytosine insertion in the VNTR of the MUC1 gene was recently identified as the most common cause of medullary cystic kidney disease type1. The phenotypic expression of MUC1 mutations have not been well characterized. Methods: Genotyping and evaluation of clinical characteristics were performed on families with a history of autosomal dominant interstitial kidney disease. Results: Twenty-five families were identified with MUC1 mutation. Of 179 family members undergoing mutational analysis, a mutation was identified in 95 individuals, and 84 individuals did not have a mutation. There were 110 individuals identified as historically affected. Individuals with a MUC1 mutation suffered from chronic kidney failure with a widely variable age of onset of end-stage kidney disease, ranging from 16 years to greater than 80 years.
3036T
Unbiased Next Generation Sequencing analysis confirms the existence of an autosomal dominant Alport syndrome. C. Fallerini1, L. Dosa1,2, D. Giachino3, R. Tito1, M. Baldassarri3, D. Del Prete4, S. Fertozzi5, G. Gai5, M. Clementi5, A. La Manna5, N. Miglietti6, R. Mancini7, G. Mandrile8, R. Artuso1, G.M. Ghiggeri10, G. Piaggio10, F. Brancati11, L. Dion11, E. Frato12, A.R. Pinciaroli13, M. Giani14, F. Castorina14, E. Bresin15, F. Mari15, M. Brutti11,2, M. De Marchi3, F. Ariani5, A. Renieri1,2. 1) Medical Genetics, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 2) San Luigi Medical Genetics, University of Turin, Turin, Italy; 3) Department of Pediatrics, University of Padova, Padova, Italy; 4) Nephrology and Dialysis Unit, Hospital Bellcolle, Viterbo, Italy; 5) S.C. Medical Genetics A.O. City of Health and Science, Turin, Italy; 6) Nephrology and Dialysis Unit, University of Padova, Padova, Italy; 7) Clinical Genetics Unit, Department of Pediatrics, University of Padova, Padova, Italy; 8) Department of Pediatrics, Second University of Napoli, Napoli, Italy; 9) Nephrology and Dialysis S.O.C. Department of General Medicine and Medical Specialties ‘Giuseppe Spada’, A.O. Pugliese-Ciaccio, Catanzaro, Italy; 14) U.O.C. PEDIATRIC NEPHROLOGY AND U.O.C. Pediatric Nephrology and Hemodialysis, Ospedale Maggiore Policlinico, Milan, Italy; 15) Clinical Research Center for Rare Diseases ‘Aldo e Cele Daccò’, Rancio, Italy.

The story of the mode of transmission of Alport syndrome (ATS) was very troubled. In 1927, the disease was hypothesized as autosomal dominant condition with male segregation distortion. In 1990, the discovery of mutations in COL4A3 or COL4A4 genes was identified and it was calculated to account for about 65% of ATS cases. A dominant form was identified most recently by the description of some large pedigrees but the real existence of this form is still questioned by many and its exact prevalence is unknown. The introduction of Next Generation Sequencing allowed us to perform an unbiased simultaneous analysis of the 3 genes (454 GS Junior, Roche) in a large cohort of Italian families (90) with clinical suspicion of ATS. In 48 of them a mutation was identified (53%). The subsequent segregation analysis in 161 family members allowed to clarify that the prevalent form was indeed the X-linked one (65%) and the autosomal recessive form (35%) was the real form. In the autosomal recessive form the protein is not truncated but its formation is reduced because of a compromised splicing of the transcripts. The autosomal dominant form must therefore be a heterozygous C-to-A mutation, resulting in a premature truncation of the protein. In both families, parental DNA showed no detectable mutation. At young ages, the boy of the first family presented with renal failure and severe hypertension, while that of the second family was diagnosed with membranoproliferative glomerulonephritis and hypertension. With time, both boys had a slow progressive deterioration in their renal function, culminating in the need for kidney transplantation. The SOX18 gene is well known to play a role in the formation of blood and lymphatic vessels, but those with homozygous mutations in this gene do not develop renal failure. Here, we report for the first time two independent cases of renal failure associated with the identical heterozygous C-to-A mutation, in a living child and his stillborn brother from Canada, as well as a living Belgian boy. The two living index cases were diagnosed with HLTS and DNA analysis for the SOX18 gene showed the same heterozygous C-to-A mutation, resulting in a premature truncation of the protein. In both families, parental DNA showed no detectable mutation. This suggests that SOX18 should be considered for the diagnosis of both ATS and HLTS. We report for the first time a case of heterozygous SOX18 mutation in a child with ATS of the recessive form, highlighting the possibility of SOX18 mutations in ATS cases with a mild phenotype. A further case of autosomal dominant ATS is also reported. The case was under investigation for a decade before the positive genetic diagnosis was obtained. The case highlights the importance of ongoing patient and family care and of performing a complete investigation, including molecular testing.
3039T
RNA-seq gene expression profiling identifies MEK1/2 and SHP2 as positive regulators of chondrocyte terminal differentiation. ME. Bowen1, UM. Ayturk1, W. Yang2, MC. Warman1. 1) Orthopaedic Research Laboratories, Children's Hospital Boston, Boston, MA 02115; 2) Department of Orthopaedics, Brown University, Providence, RI 02903.

The genetic disorder, Metachondromatosis (MC), is caused by heterozygous loss-of-function mutations in PTPN11, which encodes the phosphatase SHP2. MC patients develop benign cartilage tumors, resembling exostoses and enchondromas, that likely arise following somatic ‘second hit’ mutations in PTPN11. In a normal growth plate, chondrocytes first proliferate, then undergo terminal differentiation and are subsequently replaced by bone. To date, the molecular events in chondrocyte terminal differentiation are not well understood.

We used RNA-seq to identify genes differentially expressed after genetic inactivation of Ptpn11 in primary murine chondrocytes induced to undergo terminal differentiation in pellet cultures. Furthermore, since SHP2 has been shown to positively regulate the RAF/MEK/ERK pathway in other cell types, we treated chondrocyte pellets with U0126, an inhibitor of MEK1/2, to determine the role of this pathway in chondrocyte terminal differentiation. We found that U0126-treated pellets had increased levels of transcripts associated with proliferative, pre-hypertrophic and early-hypertrophic chondrocytes, but decreased levels of transcripts associated with late-hypertrophic chondrocytes, including multiple genes encoding proteases that degrade cartilage matrix components. Similar, but less substantial, changes in gene expression were observed after deletion of Ptpn11. We were able to identify a set of genes expression that were not dependent on MEK/ERK signaling, which were increased over time during the maturation of wild-type chondrocytes in pellet cultures. Many of these genes had not previously been associated with chondrocyte maturation, and thus represent potential novel regulators of terminal differentiation. In summary, our results suggest that SHP2 and the RAF/MEK/ERK pathway may negatively regulate the onset of hypertrophy, but positively regulate transition from pre-hypertrophic chondrocytes to terminally differentiated chondrocytes. These data are consistent with the hypothesis that enchondroma-like lesions in MC patients arise due to a failure of chondrocyte terminal differentiation.

3040F
Absence of CyPB directly affects collagen folding and glycosylation, and indirectly affects helical hydroxylation by LH1, altering bone cross-link patterns. W.A. Cabral1, I. Perdirvara2, M.A. Weis3, M. Terajima4, A.R. Blissett5, W. Chang1, E.N. Makareeva5, E.L. Mertz5, S. Leikin5, K.B. Tomer6, D.R. Eyre7, M. Yamachi8, J.C. Marini1. 1) Bone & Extracellular Matrix Branch, NICHD, NIH, Bethesda, MD; 2) Laboratory of Structural Biology, NIEHS, NIH, Research Triangle Park, NC; 3) Orthopaedic Research Laboratories, University of Washington, Seattle, WA; 4) North Carolina Oral Health Institute, University of North Carolina, Chapel Hill, NC; 5) Section on Physical Biochemistry, NICHD, NIH, Bethesda, MD.

Cyclophilin B (CyPB), encoded by Ppib, is an ER-resident peptidyl-prolyl cis-trans isomerase (PPIase) that occurs both independently and as a component of the collagen prolyl 3-hydroxylation complex. CyPB is proposed to be the isomerase catalyzing the rate-limiting step in collagen folding. Mutations in Ppib cause recessively inherited osteogenesis imperfecta type IX, a moderately severe to lethal bone dysplasia. To investigate the role of CyPB in chondrocyte terminal differentiation, we characterized a deep sequencing strategy to discover previously unknown skeletal genes.

In a normal growth plate, chondrocytes first proliferate, then undergo terminal differentiation and are subsequently replaced by bone. To date, the molecular events in chondrocyte terminal differentiation are not well understood. We used RNA-seq to identify genes differentially expressed after genetic inactivation of Ptpn11 in primary murine chondrocytes induced to undergo terminal differentiation in pellet cultures. Furthermore, since SHP2 has been shown to positively regulate the RAF/MEK/ERK pathway in other cell types, we treated chondrocyte pellets with U0126, an inhibitor of MEK1/2, to determine the role of this pathway in chondrocyte terminal differentiation. We found that U0126-treated pellets had increased levels of transcripts associated with proliferative, pre-hypertrophic and early-hypertrophic chondrocytes, but decreased levels of transcripts associated with late-hypertrophic chondrocytes, including multiple genes encoding proteases that degrade cartilage matrix components. Similar, but less substantial, changes in gene expression were observed after deletion of Ptpn11.

We were able to identify a set of genes expression that were not dependent on MEK/ERK signaling, which were increased over time during the maturation of wild-type chondrocytes in pellet cultures. Many of these genes had not previously been associated with chondrocyte maturation, and thus represent potential novel regulators of terminal differentiation. In summary, our results suggest that SHP2 and the RAF/MEK/ERK pathway may negatively regulate the onset of hypertrophy, but positively regulate transition from pre-hypertrophic chondrocytes to terminally differentiated chondrocytes. These data are consistent with the hypothesis that enchondroma-like lesions in MC patients arise due to a failure of chondrocyte terminal differentiation.

3041W
Use of whole exome sequence analysis to identify TLE4 as a causal variant in a family with congenital kyphoscoliosis and prolonged patency of the anterior fontanelle. P. Giampietro1, A. Stoddard2, D. Sweetser1, C. Raggio3, R. Blank4, M. Stephano5, K. Rasmussen6, K. Gill1, S. Sund1, R. Lorier1, A. Turner2, U. Broecke1. 1) University of Wisconsin-Madison, Madison, WI; 2) Medical College of Wisconsin, Milwaukee, WI; 3) Massachusetts General Hospital, Boston, MA; 4) Hospital for Special Surgery, New York, NY; 5) Madigan Healthcare System, Tacoma, WA; 6) Marshfield Clinic, Marshfield, WI.

Congenital vertebral malformations (CVM) represent defects in formation and segmentation of somites and have an estimated incidence of 0.13-0.50 per 1000 live births. Extreme genetic heterogeneity and the rarity of large families with CVM limit the ability to identify mutations in pattern genes associated with CVM by traditional genetic study designs. We therefore used whole exome sequencing (WES) to study a kindred in which a man and his two daughters suffered from kyphoscoliosis, vertebral body hypoplasia, short limbumar spinal pedicles, sacral and coccygeal hypoplasia and prolonged anterior fontanelle patency. WES was performed using Agilent SureSelect hybridization-based exome capture methodology. Results were filtered to exclude all except heterozygous nonconsyonymous coding variants with minor allele frequency < 1%. This strategy yielded 21 candidates, with TLE4 harboring a c.A1318G:p.T440A variant. The functional significance of the TLE4 variant was demonstrated by a mouse knockout of Tle4, the murine homolog. Tle+/- mice displayed a profound impairment of bone formation, while Tle-/- mice displayed a similar phenotype to the affected humans, including shortened vertebral pedicles and delayed mineralization of the skull. TLE4 encodes a transcription factor in a gene family that participates in the Wnt signaling pathway. TLE4 is expressed in multiple tissues, and is a critical signaling pathway in somitogenesis. The Gro/TLE family gene was not known to involved in skeletal development before the identification of TLE4 as a strong candidate gene in this family and the simultaneous characterizations of Wnt signaling and the TLE4/TLE7 family. The TLE4/TLE7 family serves as downstream effectors of Notch signaling, as well as inhibitors of Wnt/Tcf/catenin signaling. Disruption of these pathways is seen in certain families with CVMs. In addition, the Gro/TLE gene family is also known to interact with RUNX family members, which are involved in bone development. The findings highlight the potential of whole exome sequence analysis, particularly when coupled with functional studies in appropriate models, to identify novel genetic determinants of uncommon disorders. While further work remains to be done to fully define the role of TLE4 mutations in the observed phenotypes, these data nevertheless demonstrate the ability of a deep sequencing strategy to discover previously unknown skeletal genes.
Novel mutations in the LRP5 gene in patients with osteoporosis-pseudoglioma syndrome. M. Pekkinen1, G. Grigelioniene2, L. Akim2, K. Shah4, K. Karar6, S. Kurtoglu5, A.V. Ekbote3, E. Sagsak6, J. Söderholm1, S. Vallius1, S. Danda4, E. Aström3, O. Mäkitie1. 1) Folkhålska Insitut Genetik, Folkhålska Insitut Genetics, Univ Helsinki, Helsinki, University of Helsinki, Finland; 2) Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden; 3) Erciyes University, Faculty of Medicine, Department of Pediatric Endocrinology, Turkey; 4) Department of Clinical Genetics, Christian Medical College and Hospital Vellore, India; 5) Intergen, Genetic Diagnosis Research and Application Center, Ankara, Turkey; 6) Dr.Sami Ulus Children’s Hospital, Department of Pediatric Endocrinology, Ankara, Turkey.

Background: Osteoporosis-pseudoglioma syndrome (OPPG) is a rare autosomal recessive disorder with congenital or early-onset blindness and severe osteoporosis. OPPG is caused by biallelic mutations in the low-density lipoprotein receptor-related protein 5 (LRP5) gene. We present six novel LRP5 mutations and the resulting phenotypes in four consanguineous Indian and Turkish families. Methods: Peripheral blood samples were obtained from the affected probands and their parents. DNA was extracted using standard procedures. The LRP5 gene was analyzed by direct sequencing after PCR amplification. The observed sequence changes were compared with reference database and 200 control samples. Results: Altogether seven patients from three Turkish and one Indian families were included in the study; in all families the parents were related. All patients had severe osteoporosis with peripheral and vertebral fractures, and congenital or early-onset blindness. In addition, three patients had delayed mental development. DNA sequencing demonstrated in each of the four probands homozygous LRP5 mutations. All mutations were novel and located in exons 1, 2, 3 and 12 of the LRP5 gene; they were regarded deleterious based on prediction programs. In one family the affected child had homozygous sequence variations A3A, A4A, PSL and a deletion of the rs72555376 microsatellite; the healthy parents had none of these changes. In three families with altogether six affected children we found homozygous missense mutations D116N, P197R and I882N; the parents were heterozygous for the changes. Mutations P197R and I882N in exon 12 were associated with mental retardation. Conclusions: We report six novel LRP5 mutations in Turkish and Indian patients with OPPG. All mutations regardless of the location resulted in similar ocular and skeletal phenotype. Mutations in exons 3 and 12 resulted also in developmental delay. The deletion of rs72555376, which causes shortening of a poly-leucine stretch in exon 1, has previously been described in heterozygous state by Chung et al 2008; the disease mechanism and origin of the homozygous deletion together with 3 other sequence changes in our patient remains to be elucidated in further studies.
3044W
Altered Osteoblast Function underlies phenotype of type V Osteogenesis Imperfecta. A. Reich1, A.S. Bae1, A.M. Barnes1, W.A. Cabral1, D. Chitayat2, S.C. Hill3, J.C. Marini1. 1) Bone and Extracellular Matrix Branch, NICHD, NIH, Bethesda, MD; 2) The Hospital for Sick Children, and the Department of Obstetrics and Gynecology, the Prenatal Diagnosis and Medical Genetics Program, Toronto, Ontario, Canada; 3) Diagnostic Radiology Department, Clinical Center, NIH, Bethesda, MD.

Osteogenesis imperfecta (OI) is a genetically heterogeneous disorder characterized by bone fragility. Type V OI is the only OI type with dominant inheritance not caused by mutations in type I collagen; furthermore, steady-state collagen from type V osteoblasts has normal gel mobility. Affected individuals have characteristic skeletal findings, including hypertrophic callus, ossification of the interosseous membrane, as well as mesh-like lamellation on bone histology. Type V OI is caused by a unique heterozygous mutation in IFITM5 (c.-14C>T), which encodes Bril, a transmembrane protein expressed in osteoblasts. The mutation generates a start codon, adding five residues to the Bril N-terminus. However, the mechanism of type V OI and its relationship with type I collagen is unknown. We identified 8 patients with the IFITM5 (c.-14C>T) mutation, with variable phenotypic expression. Osteoblasts from type V OI patients were differentiated with osteogenic media in culture over 21 days. Sequencing of CDNA from these cells verified expression of mutant IFITM5 transcripts, consistent with a dominant mechanism. In both control and type V OI differentiated osteoblasts, Bril expression increases substantially and comparably with BMP2 stimulation. Bril protein is moderately decreased in treated mutant cells. Osteocalcin (BGLAP2) expression, a marker of late osteoblast differentiation, was increased in type V osteoblasts vs control during days 10-15 of differentiation, and relatively augmented by addition of BMP2 to cultures. Mineralization, assayed by alizarin red deposition, was also increased in type V osteoblasts during differentiation, vs control. In contrast, type V OI osteoblasts had less than half the COL1A1 expression of control during differentiation. Matrix deposited collagen when examined by confocal microscopy, but with a more sheet-like organization than in control. Finally, comparison of osteoblasts from type V OI patients to control revealed common sets of significantly altered transcripts for mineralization and growth regulation of Bril by BMP2. The increased mineralization and collagen deposition in type V OI osteoblasts during differentiation may underlie the overactive tissue calcification and hypertrophic callus formation seen in patients.

3045T
Whole Exome Sequencing and functional follow up in 114 cases of non-motile ciliopathies (Jeune-Asphyxiating Thoracic Dysplasia (JATD), Bardet-Biedl-Syndrome (BBS), Joubert-Syndrome (JS)), Short-Rib-Polydactyly-Syndrome (SRPS), Bardet-Biedl-Syndrome (BBS), Joubert-Syndrome (JS) and other ciliopathies). A.M. McInerney-Leo1, C. Cortes2, V. Plagnol3, F. Lescai4, S. Christou5, H. Jungbluth6, G. Haliloglu7, H. Kayserili8, N. Elioglu9, B. Tuyusuz10, M.E. Hurles11, P.J. Scambler1, A. Zan11,12, M.A. Brown1, E. Dincau1, U. K10, C. Wickung4, P.L. Beales1,2, H.M. Mitchison1. 1) Molecular Medicine Unit, Institute of Child Health, University College London (UCL), London, UK; 2) Centre for Translational Genomics-GOSgene, Institute of Child Health, UCL London, London, UK; 3) The University of Queensland Diamantina Institute, Translational Research Institute, Wooloongabba, Queensland, Australia; 4) Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland, Australia; 5) Department of Genetics, Environment and Evolution, UCL Genetics Institute (UGI), University College London, London, London, UK; 6) Department of Paediatric Neurology, Evelina Children’s Hospital, Guy’s and St Thomas’ National Health Service (NHS) Foundation Trust, London, UK; 7) Department of Pediatic Neurology, Hacettepe University Children’s Hospital, Ankara, Turkey; 8) Medical Genetics Department, Medical Faculty, Istanbul University, Istanbul, Turkey; 9) Department of Pediatric Genetics, Marmara University Hospital, Istanbul, Turkey; 10) Department of Pediatrics, Division of Genetics, Cerrahpasa Faculty of Medicine, Istanbul University, Istanbul, Turkey; 11) Wellcome Trust Sanger Institute, Hinxton, UK; 12) The University of Queensland, UQ Centre for Clinical Research, Herston, Queensland, Australia; 13) www.uk10k.org.

Background: Ciliopathies are complex develop-mental disorders resulting from mutations in genes encoding ciliary proteins. Although individually rare, together they represent a significant disease burden. Phenotypic as well as genetic overlap, together with extensive genetic heterogeneity, has hampered genetic diagnosis in the past but development of next generation sequencing technologies such as whole exome sequencing (WES) offers new diagnostic tools. A large non-motile ciliopathy cohort demonstrates that WES is highly efficient. Our findings in this large non-motile ciliopathy cohort demonstrate that WES is highly effective for the genetic diagnosis of heterogeneous recessive disorders and that this diagnosis is facilitated by deep pheno-typing. A surprisingly large number of cases are caused by mutations in known genes, but we also identified several new ciliopathy genes. Compared to NGS gene panel sequencing, WES offers additional opportunities to identify new genes previously not associated with the condition investigated.
Haplotype Analysis Supports a "Founder" for the Balkan OPG Mutation Causing Juvenile Paget's Disease. S. Mumm,1,2 K. Gece1, K. Huskey, D. Naot1, S. Polyzos3, T. Cundy4, W. Van Hul5, P. Singhelakis1, M.P. Whyte1,2. 1) Washington University School of Medicine, St. Louis, MO, USA; 2) Shriner's Hospital for Children, St. Louis, MO, USA; 3) University of Auckland, New Zealand; 4) Aristotle University of Thessalaki, Greece; 5) University of Antwerp, Belgium.

Juvenile Paget's disease (JPD), a rare autosomal recessive disorder features extremely rapid bone turnover causing skeletal pain, fracture, and deformity in early childhood. Deafness and retinopathy leading to blindness can follow. Most JPD is due to homozygous loss-of-function mutations in the TNFRSF11B gene encoding osteoprotegerin (OPG) -- the decoy receptor that prevents RANKL binding to its cognate receptor RANK, a major stimulus indicator for osteoclastogenesis. A variety of mutations in TNFRSF11B, typically transmitted by apparent "founders" in various geograhpic locations worldwide, cause JPD. The severity of 'OPG deficiency JPD' seems to vary depending on mutation type and how much of the OPG coding region is disrupted. In two Greek men and another Croatian man who are seemingly unrelated and manifest the mildest form of JPD, a unique homozygous deletion/insertion mutation (966_969delTACinsCTT) has been reported. This frame-shift deletes 79 carboxy terminal amino acids from the OPG monomer, including a cysteine necessary for homodimerization. When recently referred 2 additional unrelated Greek JPD patients homozygous for this 'Balkan mutation', we studied the likelihood that all of these patients are in fact related by a distant 'founder' in whom the mutation arose many generations ago. To test this hypothesis, we performed haplo- type analysis of the 5 Balkan JPD patients and 2 control JPD patients from other geographic regions using SNPs within and surrounding TNFRSF11B. We developed primer sets for 13 informative SNPs, and then PCR-amplified and sequenced these SNPs using leukocyte-derived genomic DNA from each patient. All 5 individuals shared a homozygous common core haplotype of 4 SNPs, whereas each of the 'control' JPD patients had a unique haplotype. Two of the four Greek patients shared a common haplotype of 10 homozygous SNPs, indicating closest kinship. A different Greek patient shared only 5 of these SNPs suggesting more distant relationship. The Croatian JPD patient shared only 4 common markers and four Balkan JPD patients. All Balkan patients share the common OPG mutation (966_969delTACinsCTT) and common haplotypes indicating relationship to a genetic founder. Further analysis can determine how many generations ago this founder mutation arose.

Exome analysis on tubular aggregate myopathy. Y. Endo1,2, K. Motomura1, Y.K. Hayeshti1, S. Naguchi1, I. Nonaka2, M. Mori-Yoshimura1, Y. Oya3, I. Nishino1,2. 1) Department of Clinical Development, Translational Medical Center, National Center of Neurology and Psychiatry (NCNP), Kodaira, Japan; 2) Department of Neuromuscular Research, National Institute of Neuroscience, NCNP; 3) Department of Neurology, National Center Hospital, NCNP.

Tubular aggregates (TAs) are a unique structure in muscle fibers, which is derived from sarcoplasmic reticulum, and are seen in certain types of muscle diseases including periodic paralysis, congenital myasthenic syndromes (CMS) and tubular aggregate myopathy (TAM). TAM is defined as unclassified hereditary myopathy, pathologically characterized by the presence of TAs, although causative relationship has not been well established. The objective of this study is to identify pathogenic mutations of TAM and to clarify the clinical phenotypes. We performed whole-exome sequencing in 13 individuals with TAM but with no known genetic defects, and identified pathogenic mutations in GFTP1 in two individuals. Both individuals had novel homozygous missense mutations, c.41G>A (p.Glu14Gln) and c.723_724insG (p.G241fs), respectively. GFTP1 was recently reported as a causative gene for limb-girdle CMS (LG-CMS). However, the clinical presentations of the two affected individuals were slightly different from LG-CMS. Individual 1 exhibited slowly progressive limb-girdle muscle weakness with no fluctuation therefore the clinical diagnosis was limb-girdle muscle dystrophy. Individual 2 exhibited limb-girdle and facial muscle weakness from infancy. Therefore, the clinical diagnosis was congenital myopathy. Muscle biopsy showed myopathic change with TAs although the first biopsy of individual 2 performed at age 8 years showed only minimal change without TA. Repetitive stimulation performed in individual 1 after genetic analysis demonstrated the decremental responses diagnostic of myasthenia. Acetylcholine-esterase inhibitor therapy was dramatically effective, improving his status from wheelchair-bound to ambulatory without support. GFTP1-associated LG-CMS should be considered when affected individuals have limb-girdle muscle weakness and TAs.

Novel COL11A1 mutations in Stickler syndrome detected by next-generation sequencing. F. Acket1,2, F. Maffei3, O. Vanakker4, W. Steyaert2, K. De Leeuwen1, I. Dhooge1, A. De Paepe2, E. De Leenheer1, P. Coucke1. 1) Department of Otorhinolaryngology, Ghent University Hospital, Ghent, Belgium; 2) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium.

Stickler syndrome is a connective tissue disorder with considerable phenotypic and genotypic variability, characterized by distinctive facial abnormalities, ocular problems, hearing loss, and joint problems. Up to now, mutations in 5 different collagen genes have been associated with the disease. Mutations in the COL11A1 gene may result in Stickler syndrome type 2, which can be differentiated clinically by a 'beaded' vitreous and a more severe hearing loss compared to the more prevalent type 1, caused by mutations in COL2A1. In literature, only a few dozen COL11A1 mutations have been reported. We selected 33 unrelated COL2A1 mutation-negative patients, based on their clinical features, for molecular COL11A1 analysis, by means of next-generation sequencing technology (MiSeq, Illumina). The bioinformatic pipeline included the CLC bio Workbench 6.0 followed by an in-house developed software package for variant interpretation. Assays lacking sufficient coverage as well as the mutations identified were verified by Sanger sequencing. Out of the 33 selected probands, disease-causing mutations could be identified in 14 independent patients. Half of these mutations are novel, of which most are missense mutations (including 2 glycine substitutions). The other half are previously reported mutations, mostly splice-site alterations. Moreover, additional SNPs were identified. All molecularly confirmed patients had a clinical presentation compatible with Stickler syndrome type 2. In conclusion, COL11A1 screening in selected Stickler syndromes could be a promising molecular tool.

Mutation analysis of COL1A1 and COL1A2 genes in Indian patients with Osteogenesis Imperfecta. J. Stephens1, A. Shukla2, A. Dalal2, G. Katra1, N. Gupta1, M. Kabra3, P. Dabadghao5, S. Phadke1. 1) Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Science, Lucknow, India; 2) Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; 3) Department of Medical Genetics, Kasturba Medical college, Manipal, India; 4) All India Institute of Medical Science, New Delhi, India; 5) Endocrinology, Sanjay Gandhi Post Graduate Institute of Medical Science, Lucknow, India.

Osteogenesis Imperfecta (OI) is a condition of decreased bone density of heterogeneous etiology. Most of the cases are inherited in autosomal dominant fashion and are caused by mutations in COL1A1 or COL1A2 genes. These two genes are very large so they have been rarely analyzed systematically in Indian patients with OI. We have selected 35 Indian patients who were clinically diagnosed for OI and sequenced all exons of both the genes. Mutations in COL1A1 gene were identified in 14 cases (six novel) and mutations in COL1A2 were identified in eleven cases (seven novel). Total 55 polymorphisms have been identified in both the genes; out of them eight were novel in the coding region and twelve were novel in the non coding region. No mutation has been detected in ten patients; out of them six were from consanguineous families with one or two similarly affected siblings suggesting autosomal recessive genes. Excluding consanguineous families, mutation has been identified in 25 out of 29 families which accounts for about 86 percent of mutation detection rate in these genes in Indian patients with Osteogenesis Imperfecta.
3050W

**WNT7** Mutations in Early-onset Osteoporosis and Osteogenesis Imperfecta. P. Carpeau1, C. Laine1,2, K.S. Koeng1, R. Kiviranta2,3, K. Tarkkonen4, M. Grover1, J.T. Lu1, M. Pekkinen5, M. Weissman2,6, T.J. Heino1, V. Nieminen-Pihala1, M. Aronen1, T. Laine1,2, H. Kröger1, W.G. Cole1, A.E. Lehbesioki1,2,3,4, L. Nevaraz5, D. Krakow6,7, C.J.R. Curry8,9, D.H. Cohn10,11, R.A. Gibbie12,13, O. Mäkitie14,15, B.H. Lee16,17,1. Molec & Human Gen, Baylor College Med, Houston, TX, USA; 2) Folkvålan Institute of Genetics, Helsinki, FINLAND; 3) Department of Endocrinology, Sahlgrenska University Hospital, Gothenburg, SWEDEN; 4) Department of Medical Biochemistry and Genetics and Department of Medicine, University ofTurku, Turku, FINLAND; 5) Department of Medicine, Turku University Hospital, Turku, FINLAND; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 7) Department of Structural and Computational Biological & Molecular Biophysics, Baylor College of Medicine, Houston, TX, USA; 8) Institute of Molecular Medicine Finland, University of Helsinki, Helsinki, FINLAND; 9) Department of Cell Biology and Anatomy, University of Turku, Turku, FINLAND; 10) Department of Pediatric Orthopedic Surgery, Sahlgrenska University Hospital, Gothenburg, SWEDEN; 11) Bone and Cartilage Research Unit, University of Eastern Finland and Kuopio University Hospital, Kuopio, FINLAND; 12) Division of Pediatric Surgery, University of Alberta, Edmonton, CANADA; 13) Haartman Institute, Department of Medical Genetics and Research Program's Unit, Molecular Medicine, University of Helsinki, Helsinki, FINLAND; 15) Department of Molecular, Cell and Developmental Biology, University of California-Los Angeles, CA, USA; 16) Department of Orthopaedic Surgery, University of California-Los Angeles, CA, USA; 17) Department of Human Genetics, University of California-Los Angeles (DKFZ), Heidelberg, Germany; 18) University of California-San Francisco, San Francisco, CA, USA; 19) Children's Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, FINLAND; 20) Howard Hughes Medical Institute, Houston, TX, USA.

This report identifies human skeletal diseases associated with mutations in **WNT7**. In ten family members with dominantly inherited early-onset osteoporosis, a heterozygous missense variant c.652T>G (p.Cys218Gly) in **WNT7** segregated with the disease, and a homozygous nonsense mutation (c.884C>A, p.Ser295*) was identified in two siblings with recessive osteogenesis imperfecta. In vitro, aberrant forms of **WNT7** protein showed impaired capacity to induce canonical WNT signaling, their target genes, and mineralization. **WNT7** was clearly expressed in bone marrow, especially in B cell lineage and hematopoietic progenitors; lineage tracing identified expression in a subset of osteocytes, suggesting altered cross-talk of WNT signaling between hematopoietic and osteoblastic lineage cells in these diseases.

3051T

**Identification of novel SHOX target genes in the developing limb using a transgenic mouse model.** G.A. Rappold1, A. Glaser1, K. Kleinschmidt1, I. Scholl1, R. Röth2, L. Li3, N. Grettz4, G. Mechtersheimer1, M. Karperien4, A. Marchini5, W. Richter5, K.U. Beiser1. 1) Department of Human Molecular Genetics, Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany; 2) Division of Experimental Orthopaedics, Orthopaedic University Hospital, Heidelberg, Germany; 3) Institute of Pathology, University of Göttingen, Göttingen, Germany; 4) Medical Research Center (ZMF), Medical Faculty Mannheim at Heidelberg University, Mannheim, Germany; 5) Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany; 6) Department of Developmental Bioengineering, University of Twente, Enschede, The Netherlands; 7) German Cancer Research Center (DKFZ), Heidelberg, Germany.

Deficiency of the human short stature homeobox-containing gene (**SHOX**) has been identified in several disorders characterized by reduced height and skeletal anomalies such as Turner, Léri-Weil and Langer syndrome as well as idiothetic short stature. **SHOX** acts as a transcription factor during limb development and is expressed mainly in prehypertrophic and hypertrophic chondrocytes of the growth plates. Although highly conserved in vertebrates, rodents lack a **SHOX** ortholog. This offers the unique opportunity to analyze the effects of human **SHOX** expression in transgenic mice. We have generated a mouse expressing the human **SHOX** cDNA under the control of a murine Col2a1 promoter and enhancer. **SHOX** and marker gene expression as well as skeletal phenotypes were characterized in two transgenic lines. No significant skeletal anomalies were found in transgenic compared to wildtype mice. Quantitative and in situ hybridization analyses revealed that the Col2a1-**SHOX** transgene, however, affected extracellular matrix gene expression during early limb development, suggesting a role for **SHOX** in growth plate assembly and extracellular matrix composition during long bone development. For instance, we could show that the connective tissue growth factor gene (**Ctgf**), a gene involved in chondrogenic and angiogenic differentiation, among other genes is transcriptionally regulated by **SHOX** in transgenic mice. This finding was confirmed in human NHDF and U2OS cells and chicken micromass culture, demonstrating the value of the **SHOX**-transgenic mouse for the characterization of **SHOX**-dependent genes and pathways in early limb development.
3053W

Discovery of a novel genetic disorder and increased diagnostic rate using Next Generation Sequencing in heterogeneous Ataxias. A.H. Ragoussis, L. Tsiora, C. Kwasniewska, S. Liss, R. Faroli, Schnekenberg, E.B.E. Becker, K.D. Bera, M.E. Shanks, L. Gregory, D. Buck, M.Z. Cader, K. Talbott, R. de Silva, G. Fletcher, R. Hastings, S. Jayaward, P.J. Morrison, P. Worth, M. Taylor, J. Tomle, M. O'Regan, R. Valentine, E. Packham, J. Evans, P. Clouston, A. Sellor, J. Ragoussis, UK Ataxia Consortium. 1) Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom; 2) Department of Clinical Genetics, Churchill Hospital, Oxford University Hospitals NHS Trust, Oxford, OX3 7LD; 3) Welcome Trust Centre for Human Genetics, University of Oxford, OX3 7BN; 4) School of Medicine, Universidade Positiva, Curitiba, Brazil; 5) Department of Physiology, Anatomy and Genetics, MRC Functional Genomics Unit, University of Oxford, OX1 3QX; 6) Department of Neurology, Essex Centre for Neurological Sciences, Queen's Hospital, Romford; 7) Walton Centre NHS Foundation Trust, Liverpool, L9 7LJ; 8) Department of Clinical Genetics, St Michael’s Hospital, Bristol, BS2 8EG; 9) Department of Paediatrics, Oxford University Hospitals NHS Trust, Oxford, OX3 7LJ; 10) School of Medicine, Dentistry and Biomedical Sciences, Queens University, Belfast; 11) Department of Neurology, Norfolk and Norwich University Hospital, Norwich; 12) School of Cancer Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT; 13) Department of Clinical Genetics, Southern General Hospital, Glasgow G51 4TF; 14) Fraser of Allander Neurosciences Unit, Royal Hospital for Sick Children, Glasgow G3 8SJ; 15) Thames Valley Dementia and Neuromuscular Diseases Network, Oxford; 16) Oxford Regional Molecular Genetics Laboratories, Oxford University Hospitals NHS Trust; 17) McGill University and Génome Québec Innovation Centre, McGill University, Montreal, Canada.

Ataxias are highly heterogeneous disorders caused by mutations in many genes. The diagnostic process is often long and many patients undergo multiple investigations without ever reaching a conclusive molecular diagnosis. The diagnosis is often made relatively late, parallel to the loss of multiple multiple cochlea cell types. Functional studies are now in progress to confirm the pathogenic effect of the variation identified. As regards the second point, a Targeted Re-Sequencing (TRS) protocol has been developed based on Ion Torrent™ technology. It analyzes 96 HHL related genes ensuring correct selection of 12 samples. The protocol has been tested and validated in many cases and is currently used in our diagnostic service. These findings definitely increase our knowledge of new HHL genes and may suggest new targets for hearing impairment treatment and prevention.

3054T

Shouting won’t help: a combined strategy (Next Generation Sequencing and Linkage analysis) to identify Hereditary Hearing Loss (HHL) genes in affected families. D. Vosci, G. Giroto, F. Faletra, E. Rubrato, A. Morgan, D. Vuckovic, S. Lenarduzzi, M. Morgutti, P. Gasparini. 1Med Genet, IRCCS-Burlo Garofolo Children Hospital, Via dell'Istria 65/1, Trieste, Italy.

Non-syndromic HHL is a common disorder accounting for at least 60% of prelingual deafness. Despite the presence of some common HHL genes (GJB2, GJB6 and MTRNR1) still there is a need to 1) search for new causative mutations/genes and 2) develop new diagnostic approaches to overcome the problem of genetic heterogeneity. As regards to the first point a combined strategy based on linkage analysis (LA) followed by whole-exome sequencing (WES) was developed. Data analysis pipeline is based on data quality evaluation, hg19 reference reads mapping by BWA, variants calling and quality filtering by GATK. ‘In silico’ analysis of variants was done using a series of tools (PolyPhen-2, MutationTester, etc.). Using this approach we analyzed 6 Italian and 5 Qatari families affected by sensorineural, moderate to severe HHL. In a consanguineous family from Qatar, LA identified a region of 40 Mb on chromosome 15q13 (LOD score 3.8). WES data reported a causative mutation (c.1875+5G>A) leading to p.?-265Gluext*11) in BD1 gene. The mutation disrupts the termination codon of the transcript resulting in an elongation of 11 residues of the protein. Immunohistochemistry in the mouse inner ear showed a clear expression of Bdp1 gene. In an Italian dominant family, LA showed a LOD of 3.3 on chromosome 12p24, where WES data analysis identified a new missense mutation (c.1057G>C; p.G353R) in P2X2 gene. A 3D model of this protein suggests that the substitution of the hydrophobic Gly, with a charged residue as the Arg, it is expected to destabilize the protein folding. Finally, an Italian pedigree resembling a model neurogenetic disorder to assess the introduction of NGS into the diagnostic service. These findings definitely increase our knowledge of new HHL genes and may suggest new targets for hearing impairment treatment and prevention.

3055F

Genetic etiology study of the non-syndromic deafness in Chinese Hans by targeted next-generation sequencing. T. Yang, Y. Li, H. Wu, Y. Chai, A. Morgan, D. Vuckovic, S. Lenarduzzi, M. Morgutti, P. Gasparini. 1Med Genetics, IRCCS-Burlo Garofolo Children Hospital, Trieste, Italy.

Although over 60 non-syndromic deafness genes have been identified to date, the etiologic contribution of most deafness genes remained elusive. In this study, we addressed this issue by targeted next-generation sequencing of a large cohort of non-syndromic deaf probands. Probands with mutations in common screened deafness genes (GJB2, SLC26A4 and MT-RNR1) were pre-excluded by Sanger sequencing. The remaining 125 deaf probands proceeded through targeted exon capturing of 79 known deafness genes and Illumina HiSeq2000 sequencing. Bi-allelic mutations in 15 less commonly screened deafness genes were identified in 28 deaf probands, with mutations in MYO15A, GPR98, TMC1, USH2A and PCDH15 being relatively more frequent (≥23 probands each). Dominant mutations in MYO6, TECTA, POUSF3 and COCH were identified in 4 deaf families. A mitochondrial MTT1 mutation was identified in one maternally inherited deaf family. No pathogenic mutations were identified in these three dominant deaf families and two consanguineous families. We concluded that mutations in the less commonly screened deafness genes were heterogeneous and contributed to a significant percentage (17.4%) of causes for non-syndromic deafness. Targeted next-generation sequencing provided a comprehensive and efficient diagnosis for known deafness genes. Complementary to linkage analysis or whole-exome sequencing of deaf families, pre-exclusion of known deafness genes by this strategy may facilitate the discovery of novel deafness genes.
3056W

Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex III deficiency. S.M.K. Farhan1,2, J.L. Den1, J.P. Robinson1, P. Lahiry1, V.M. Su1,3, C. Prasad1,5,6, J.B. Kronick1, D.A. Ramsay2, C.A. Rupar1,6, R.A. Hegele1. 1) Department of Biochemistry, Western University, London, Ontario, Canada; 2) Robarts Research Institute, Western University, London, Ontario, Canada; 3) Department of Pathology, London Health Sciences Centre, University of Western Ontario, London, Ontario; 4) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, Ontario; 5) Medical Genetics Program, Department of Pediatrics, London Health Sciences Centre, London, Ontario; 6) Children’s Health Research Institute, London Ontario.

Iron-sulfur (Fe-S) clusters are a class of highly conserved and ubiquitous prosthetic groups with unique chemical properties that allow the proteins that contain them. Fe-S proteins, to assist in various key biochemical pathways. Mutations in Fe-S proteins often disrupt Fe-S cluster assembly leading to a spectrum of severe disorders such as Friedreich’s ataxia or ISCU myopathy. Herein, we describe infantile mitochondrial complex III deficiency, a novel autosomal recessive mitochondrial disease characterized by lactic acidemia, hypotonia, respiratory chain complex II and III deficiency, multisystem organ failure and abnormal mitochondria. We applied autozygosity mapping to identify a homozygous region on chromosome 20 containing 453 genes. Next through whole-exome sequencing, we identified a novel variant within the autozygous region: c.215G>A, p.Arg72Gln in NFS1, a highly conserved cysteine desulfurase involved in the Fe-S cluster assembly machinery and is essential for the maturation of other Fe-S proteins such as ISCU, ISD11 and FXN. NFS1 p.Arg72Gln co-segregates with disease status in the family and was consistently predicted to be damaging by multiple in silico analyses. We thus describe the first disease in man likely caused by deficiency in NFS1. Our results further demonstrate the importance of NFS1 expression in human physiology.

3057T

Reducing the cost of whole-exome sequencing of parent-affected offspring trios by joint Bayesian variant calling. B. Hilbus1, J. Blue-Smith1, S. Lombardi1, R. Litten1, L. Trigg2, A. Jackson2, D. Ware2, J.G. Cleary2. 1) Real Time Genomics, Inc., San Bruno, CA; 2) Real Time Genomics, Inc., Hamilton, NL.

Whole-exome sequencing (WES) has been a successful and cost-effective strategy to identify disease-causing mutations of Mendelian highly penetrant diseases in research and clinical settings. The power to identify disease mutations is much higher by sequencing parent-affected offspring trios rather than the proband alone as genotypes of the parents remove irrelevant variants, aids in the removal of false positives shown as Mendelian inheritance errors (MIEs), and permits identifying recessive patterns. However, the cost of trio-based sequencing has limited its adoption in both clinical and research settings resulting in unsolvable cases. To circumvent this barrier, here we demonstrate a novel informatics approach to substantially reduce the cost of WES of trios leveraging the shared pedigree data allowing for full coverage sequencing of the proband and half coverage of the parents while maintaining high sensitivity and accuracy. Key to our results is the use of an autozygous region as a virtual control to map the alignment data of all family members when examining a site, scoring genotypes with priors based on Mendelian segregation patterns. To validate the approach, we obtained sequence data for a CEPH trio (Illiunma 100bp paired-end, Agilent SureSelect v4 UTR) at a sequencing depth of ~60X coverage for NA12878 and at both 60X coverage and 30X coverage for the parents. We called variants on the full and half parental coverage trio and analyzed sensitivity and call set metrics for each of the individuals. We found that even with reduced parental coverage, we maintained 97% sensitivity to detect in NA12878, with 78% and 65% sensitivity for the 60X and 30X of parent coverage, respectively. We then applied this approach on OMNI data in the targeted regions, and could call 98.6% of the offspring variant sites in both parents. Genotype concordance between full and half coverage for the child was 99.79% while both parents were above 99.4%. We also detected 78 200 bp de novo variants, of which 65 were novel disease likelihood candidate mutations, but no de novo or copy number variant (CNV). In the 100X lower than that obtained with a singleton caller for NA12878. Our results show that our approach allows for a rational allocation of sequencing capacity in pedigree studies, resulting in cost savings with little deterioration of call quality. In a panel of 17 parent-affected trios, this savings effectively result in sequencing three exomes for the cost of two, greatly empowering disease and clinical studies.

3058F

Systematic identification of causal mutations in Mendelian disorders using exome sequence data. B. Thomas1,2, M. Lek1,2, N. Clarke2, L. Waddell2, C. Cardona2, D. Daly3, B. Bonnemann1,6, K.N. North5, D.G. MacArthur1,3,4. 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 2) Broad Institute of Harvard and MIT, Boston, MA, USA; 3) Institute for Neuroscience and Muscle Research, Sydney, NSW, Australia; 4) Department of Translational Medicine, Institut de Genétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U964, CNRS UMR7104, Université de Strasbourg, Collège de France, 67404 Illkirch, France; 5) Department of Paediatrics, University of Manchester, Manchester, M13 9PL, UK; 6) Neuro-muscular and Neurogenetic Disorders of Childhood Section, NINDS, National Institutes of Health, Bethesda, MD, USA.

Exome sequencing has proven to be a powerful and cost-effective approach to identifying the mutations in many families suffering from rare, severe Mendelian diseases. However, exome analysis unambiguously identifies a causal mutation in only 30-50% of sequenced families, indicating much work remains to be done to increase the yield of causal variants from sequencing-based approaches. Causal mutations can be missed by current exome sequencing approaches for a variety of reasons. Conversely, there may be multiple gene candidates that require further information to prioritize for functional studies. We describe the development of an integrated pipeline of data filtration, causal variants filtering and its application to a cohort of severe, undiagnosed muscle disease patients. Our online application called xBrowse enables the intuitive analysis of family-based exome data, permitting researchers and clinicians to rapidly explore the effects of altering inheritance modes and function/quality filters on the identified set of potential causal candidates. Moreover, our collaborators have early access to gene-based RNA expression data across various human tissues, a disease gene-centric protein interaction networks and a large reference panel of over 50,000 exomes. We have applied this integrated approach to 250 individual families with severe muscle weakness and probands affected by a range of neuromuscular diseases. We describe the detection of novel sequence variants with strong evidence for causality in these patients, and provide case studies indicating the value of tissue expression data, protein interaction networks, large reference panels for the prioritization of disease-associated mutations.
3060T

Mild case of Unverricht-Lundborg disease presenting as juvenile myoclonic epilepsy. E. Andermann1,3,4,5,6,1, D. Amrom1,2,3,4,6,1, A.-L. Lehesjoki1,2, 1) Neurogenetics Unit; 2) Epilepsy Service and Seizure Clinic; 3) Montreal Neurological Hospital and Institute, Montreal, Quebec, Canada; 4) Department of Neurology & Neurosurgery; 5) Department of Human Genetics; 6) Department of Pediatrics; 7) McGill University, Montreal, Quebec, Canada; 8) Folkalsan Institute of Genetics and Neurosciences Centre, University of Helsinki.

Unverricht-Lundborg disease (ULD) (EPM1) is a progressive myoclonus epilepsy with variable severity and course, and variable degrees of cognitive deterioration. The disease usually begins between 6 and 15 years of age with myoclonus and generalized tonic-clonic seizures. With good anti-epileptic management, the patients can now survive into the 50's and 60's. The gene for this disorder was identified in 1996 as cystatin B (CSTB), a cysteine protease inhibitor. The most common mutation is a dodecamer repeat, although rare point mutations have also been described. A 30-year-old female patient had one generalized tonic-clonic seizure during sleep at the age of 11 years, and onset of myoclonic jerks on awakening at around the same time, which are well-controlled with valproic acid. She carries a clinical diagnosis of juvenile myoclonic epilepsy (JME). The patient was born prematurely at 25 weeks’ gestation weighing 750 grams. Her developmental milestones were normal and she works as a highschool teacher for behaviourally challenged children. The parents were not known to be consanguineous, although they originated from two small towns 35 kms apart. Four siblings of the maternal grandmother were diagnosed clinically with ULD and were known to us; three sisters died in their 20’s and 30’s, and one brother died at age 65. A third degree cousin of the mother had epilepsy and died at 18 years of age. Her father had a single generalized tonic-clonic seizure at age 65. A third degree cousin of the mother had epilepsy and died at 18 years of age. The male patient had a case of myoclonic seizures at the age of 7 years. A distant cousin of the paternal grandmother was also said to have progressive myoclonic epilepsy. The patient and her partner presented for preconception genetic counseling. Carrier screening for the CSTB gene was carried out for both the patient and her partner, employing PCR for the dodecamer repeat, and sequencing of the CSTB gene to rule out point mutations. CSTB testing in the patient surprisingly revealed that she was a compound heterozygote for two mutations: an expansion of the dodecamer repeat and a splice site c.67-1G>C mutation in intron 1, predicting a deletion of the downstream exon 2 with in-frame expansion of the dodecamer repeat and a splice site c.67-1G>C mutation.

3060W

Next Generation Sequencing Defines New Gene(s) Involved in the Enlarged Vestibular Aqueducts and Pendred Syndrome. F. Alasti, M. Hildebrand, T. Yang, R.J. Smith. Molecular Otolaryngology and Renal Research Laboratories, Department of Otolaryngology, University of Iowa, Iowa City, 52252 IA. USA.

In patients with sensorineural hearing loss (SNHL), enlarged vestibular aqueduct (EVA), cochlear hypoplasia and Mondini (both EVA and cochlear hypoplasia) have been considered as the most common radiological finding. Pendred Syndrome (PS) is characterized by SNHL, temporal bone anomalies, and thyroid goiter. It is believed that non-syndromic DFNB4 SNHL, EVA, Mondini dysplasia and PS are parts of the same disease spectrum. Biallelic mutations in SLC26A4, is the main cause of the PS-Mondini-EVA-DFNB4 disease spectrum. Pendred, the anion exchanger encoded by SLC26A4, likely helps maintaining the fluid level and ion composition of endolymph, which is highly important during development of the shape of the bony structures of inner ear. However, mutations in cis- and trans-regulatory elements of SLC26A4 (transcriptional regulator FOX11) as well as the potassium channel KCN110 have also been reported in carriers of SLC26A4 mutations. Our goal is to study more about the possible physiological mechanisms of ionic homeostasis of the endolymph. We believe that identification of new gene(s) involved in this disease can open new insights into the pathophysiological mechanisms of SLC26A4-related SNHL. We hypothesized that using targeted exome capture followed by massively parallel sequencing on our targeted of PS-Mondini-EVA-DFNB4 patients carrying 0 or 1 mutation in SLC26A4; we can likely identify additional gene(s) associated with this phenotype. To address this issue, we designed RNA baits and developed a platform (called EVASeq) by targeting the coding exons of the best candidate genes extracted from our previous murine microarray expression data. Our strategy was screening our samples for the EVAseq platform, as well as for the OtoSCOPE (a platform targeting all the known HL-causing genes) panel. Then, the pedigrees with multiple affected individuals, in which all the genes in both EVAseq and OtoSCOPE panels were excluded, were considered for Whole Exome Sequencing (WES). Sequencing the candidate gene(s) we found through WES in a large cohort of patients with the same clinical data (in order to find a second family), was a method of screening for the relevant function and possible ways for the gene(s) we found within our WES families is our current approach. The result of this study most likely will provide the first comprehensive insight into the identification of new genetic components involved in this PS-Mondini-EVA-DFNB4 disease spectrum.

3061F

Whole exome sequencing combined with homozygosity mapping in a family with mental retardation, muscle weakness, and abnormal movement. E. Jaberi1, B. Farham2, G.A. Shahidi3, M. Rohani1, I. Safarian2, B. Klotzle3, E. Elahi1,5, 1) School of Biology, College of Science, University of Tehran, Tehran, Iran; 2) Iran University of Science and Technology, Tehran, Iran; 3) Dept. of Neurology, Tehran University of Medical Sciences, Tehran, Iran; 4) Immunia, San Diego, California, USA; 5) Dept. of Biotechnology, College of Science, University of Tehran, Tehran, Iran.

We performed homozygosity mapping using a consanguineous Iranian family with three affected and five unaffected individuals with mental retardation, muscle weakness, and abnormal movement using high density SNP chips. The pattern of inheritance was autosomal recessive. Their symptoms started from childhood with learning difficulty and evaluate as moderate mentally retarded. In addition to mental retardation and muscle weakness, they had severe kyphoscoliosis and underwent spinal surgery. All patients had normal sensory and cerebellar examination. Brain MRIs revealed mild cerebellar, cerebellar and brainstem atrophy within the affected family. Electromyography and nerve conduction studies showed normal sensory and motor action potentials but neurogenic MUAPs in distal and upper extremity muscles. Disease status in the family linked to a homozygous region of 13 Mb on chromosome 6. We captured exomes of one affected individual from the family and performed sequencing analysis by a second-generation sequencer with a mean coverage of 30x and sufficient depth to call variants at ~97% of each targeted exome. The genetic variants in the homozygous region were filtered against the 1000 Genomes Project, HapMap, and the dbSNP131 database. After annotation and functional expectation, one gene with a novel splice site disruption was found to be candidates for the disease segregated in the family. The candidate gene has not been reported for any disease before. The observed variation is a 3193G>A, p.V10645I; c.96818G>A, p.R32273H variant, which is not seen in 400 control individuals. According to the sequences of cDNAs of the 3 affected individuals compared to controls, the mutation causes early stop codon which results in synthesis of a truncated protein. Further functional studies are needed to validate the relation of the candidate gene with the disease phenotype in the family.

3063T

Haploinsufficiency of GJB5 identified via exome sequencing causes a novel form of cutis laxa. M. Dasouki1,2, J. Roberts3, K. Gonzalez4, W. Zeng3, M. Butler4, A. Belousov5, I. Saadi6, 1) Dept. of Neurology, Univ Kansas Med Ctr, Kansas City, KS; 2) Dept. of Genetics, King Faisal Specialist Hospitals & Research Center, Riyadh, Saudi Arabia; 3) Dept of Psychiatry, Univ Kansas Med Ctr, Kansas City, KS; 4) Ambry Genetics, Aliso Viejo, CA; 5) Dept. of Anatomy & Cell Biology, Univ Kansas Med Ctr, Kansas City, KS.

Cutis laxa is a clinically and molecularly heterogeneous genodermatosis. Mutations in five connexin genes (GJA1, GJB2, GJB3, GJB4, and GJB6) have been associated with a range of heritable skin disorders. Human GJB5, GJB4 and GJA4 gene cluster maps within a 55 kb genomic region at chr.1p34.3. GJB5 is a novel gene, which is highly expressed in skin and placenta. Gjb5 null mice have reduced viability. So far, two large microdeletions encompassing GJB5 has been reported. A 4.8 Mb deletion was found in a patient with hypothyroidism, low birth weight, short neck, microtia, micrognathia, strabismus and speech delay while a 2.1 Mb microdeletion was reported in an unrelated child with mental retardation. We ascertained 2 unrelated Iranian pedigrees who originally presented with chronic arthralgia, osteopenia, cutis laxa and heart palpitations. Routine clinical evaluations showed persistent large joint hypermobility, significant cutis laxa and left ventricular dysfunction. She had several unsuccessful surgical corrections of her lax skin. Magnetic resonance angiography of her chest was normal. She was treated with beta blockade and intravenous bisphosphonate. Exome sequencing of her cDNA revealed a deleterious frame shift mutation [c.37delG; p.V13fsx26] in GJB5. Two unrelated additional variants [c.3193G>A, p.V10645I; c.96818G>A, p.R32273H] were identified in TTN (encoding Pendrin, the anion exchanger) and in TNI. These TNI variants provide a plausible explanation for the ventricular dysfunction in our patient. However, this is the first report of an apparently deleterious mutation in GJB5 causing cutis laxa. Functional studies using Hela cell scratch assays and patch clamp analyses are underway to characterize the GJB5 mutation.
3064F
Identification of mutations causing congenital anomalies of the kidney and urinary tract through targeted sequencing. N. Nicolau1, J.J. Nijman1, A.M. van Eerde2, G. Groenen3, E.M. Boogaard, E. Cuppen1, K.Y. Renkema1, N.V. Knoers1. 1) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands; 2) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Congenital anomalies of the kidney and urinary tract (CAKUT) are developmental disorders that involve a spectrum of renal structural malformations. They occur in 1 of 500 live-births and comprise the major cause of end-stage renal disease. There is another category of CAKUT which is the de novo gene mutations in two genes with known heterogeneous and dominant spinal muscular atrophy. In conclusion, we discovered two heterozygous, single base change from G to A DYNC1H1 mutations in MYL3 that were further validated by Sanger sequencing. Here we show that our targeted sequencing approach and variant prioritization method is efficient in identifying gene mutations in a large cohort of sporadic CAKUT cases. Previously reported causal mutations in known CAKUT genes were successfully identified in our cohort. Interestingly, the majority of the novel and promising pathogenic variants that we identified were unique for each patient. Hence, CAKUT might be even more heterogeneous in their etiology than expected. Functional studies to test the impact of novel mutations on protein function and kidney development are currently on-going.

3065W
De novo germline mutations in ZNF259 and other genes discovered by exome sequencing in an infant with congenital joint contractures. D.W. Sant1, R.L. Margraf1, J. Durtshich1, T.M. Newcomb1, J.M. Opite2,4,5,6, J.C. Carey3, H. Zhou1, B.E. Katz2, K.V. Voelkeling1, K.J. Swoboda3,6. 1) ARUP Laboratories, Salt Lake City, UT; 2) Department of Pathology, University of Utah, Salt Lake City, UT; 3) Utah Neurology, University of Utah, Salt Lake City, UT; 4) Department of Human Genetics, University of Utah, Salt Lake City, UT; 5) Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT; 6) Department of Pediatrics (Division of Medical Genetics), University of Utah, Salt Lake City, UT. This report presents a family with one daughter affected by a previously unclassified disease, characterized by congenital joint contractures. Parents and four siblings are unaffected. To identify the potential causal variants, exome sequencing was performed on the proband, one sister, and both parents. A heuristic filtering approach identified several rare mutations, yielding the de novo mutations in the genes MYL3 and DYNC1H1 that were potentially correlated with the disorder and presented in this family. Further sequencing showed that the mutations were both present in the affected daughter, but not in either of the parents or siblings, showing that both mutations arose de novo. The de novo mutation in MYL3 is a heterozygous, single base change from A to T on position chr3:46699880. This variant causes a single-base amino acid change from tyrosine to asparagine (p.Y185N). This base is highly conserved between humans and mice to zebrafish and fruitfly. MYL3 encodes the myosin light chain 3, an alkali light chain that is necessary for contraction of cardiac muscle and slow skeletal muscle. Mutations in this gene result in muscular dystrophy and have been associated with hypertrophic cardiomyopathy, Duchenne muscular dystrophy, and familial hypertrophic cardiomyopathy-8. The de novo mutation in DYNC1H1 is a heterozygous, single base change from G to A at position chr14:102446717. This variant causes a single-base amino acid change from arginine to glutamine (p.R624Q). This base is highly conserved from humans to fungi. DYNC1H1 codes for a cytoplasmic dynein protein that is important for cell movement and ATPase activity. Mutations in DYNC1H1 are associated with a dominantly inherited autosomal spinal muscular atrophy. In conclusion, we discovered two de novo heterozygous mutations in two genes with known heterogeneous and dominant phenotypes. Future work will determine which of these two mutations is likely causative or if both are contributing to cause this novel disorder.

3066T
ZNF259 is a candidate gene for alopecia-primordial dwarfism-renal syndrome. A.G. Smith1, P. Gelcich2, A. Ahmed3, S.L. Sawyer1, D.E. Bulman1, K.M. Boycott, FORGE Canada Consortium. 1) Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) University of New Mexico Health Science Center, Albuquerque, NM. Alopecia-primordial dwarfism-renal syndrome (APDRS) is a distinctive pattern of malformation characterized by congenital alopecia, pre- and postnatal dwarfism and hypoplastic kidneys. This syndrome was observed in 5 Hispanic New Mexican children from four families. All had symmetric intrauterine growth restriction (≤3 SD below the mean) and progressively severe postnatal growth failure (≥8 SD below the mean). Skin color is very pale with a porcelian-like appearance. The craniofacial appearance is distinctive with prominent forehead, deep-set and up-sloping eyes, small nose, downturned corners of the mouth and full cheeks. One child had cleft palate, two had shunted hydrocephalus and both males had genital hypoplasia. All had severe global delays and sensorinerveal hearing loss. Four children died at ages 10, 12, 20 and 31 months, from uremia and/or sepsis. The surviving patient is a 16 yo girl with renal failure, an unusual central fat distribution and type 1 diabetes. Whole exome sequencing performed on the surviving patient and her unaffected parents yielded a homozygous missense mutation in the patient at a highly conserved locus of the ZNF259 gene, both her parents were carriers. Sanger sequencing of unaffected parents and siblings of a deceased patient demonstrated that these patients and one sibling were also heterozygous carriers of the mutation, although no patient DNA is available. ZNF259 encodes a zinc finger protein with diverse cellular localization and functions. It binds to the inactive form of the EGFR receptor in quiescent cells and is then released upon activation. It has been demonstrated to bind to eukaryotic translation elongation factor 1A in proliferating cells, form a multiprotein complex with SMN protein and accumulate in subnuclear structures. Reduction of ZNF259 expression in mammalian cells by siRNA knockdown causes widespread phenotypes, including defects in transcription, prevention of DNA synthesis and accumulation of cells in the G1 and G2 phases of cell cycle. It has been hypothesized the profound inhibition of transcription and cell cycle arrest is due to a defect in pre-mRNA splicing, since cells lacking ZNF259 have been shown to have a defective splicing machinery. ZNF259 is strongly expressed on patient fibroblast cells, as well as validating the mutation in other patients presenting with clinical features of APDRS are needed to confirm the role of ZNF259 in the pathogenesis of the disorder.

3067F
Novel MSX1 Mutations in Japanese Tooth Agenesis Patients. S. Yamaguchi1,2, J. Machida2, M. Kaminamoto2, M. Kinuma2, A. Shibata2,5,2, T. Tatematsu3,4, H. Miyashii2, Y. Higashi2, P. Jezewski5, A. Nakayama2, K. Shimozato2, Y. Tokita2,1. 1) Department of Dentistry and Oral Surgery, Aichi Children’s Health and Medical Center, Obu, Aichi, Japan; 2) Department of Maxillofacial Surgery, School of Dentistry, Aichi-Gakui University, Nagoya, Japan; 3) Department of Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Department of Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Toyota, Aichi, Japan; 4) Japanese Red Cross Society Himeji Hospital, Himeji, Japan; 5) Department of Perinatology, Institute for Developmental Research, Aichi Human Health Service Center, Kasugai, Japan; 6) Department of Periodontology, University of Alabama at Birmingham School of Dentistry, Institute of Oral Health Research, Birmingham, USA; 7) Department of Embryology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan. Since MSX1 and PAX9 are linked to the pathogenesis of non syndromic oligodonta and hypodontia, we performed a detailed mutation analysis of the MSX1 and PAX9 genes sampled from Japanese tooth agenesis cases. We identified two novel MSX1 missense mutations with an amino acid substitution in homeodomain, Thr174Ile(T174I) from a sporadic hypodontia patient from Japan and Leu205Met(L205M) from a familial oligodontia case. Both Thr174Ile and Leu205Met in MSX1 are highly conserved amino acids with in homeodomain among wide range of species. To define a possible role of these variants in the pathogenesis, we performed several functional analyses; Western blotting, immunocytochemistry, luciferase reporter assay and electrophoretic mobility shift assay (EMSA). Although the gene products are stable and capable of normal nuclear localization in transfected cells, these variants lose suppression activity on myod-promoter in cells with differentiated condition. As the result of EMSA, we clarified that the DNA binding ability of both mutated protein were reduced. On the basis of the above result, we favor of a dominant-negative effect of the Thr174Ile and L205Met as causative role of the T174I and L205M in the MSX1 gene in tooth agenesis and suggest that these amino acid substitutions at the homeodomain of MSX1 may influence cell proliferation and differentiation resulting into tooth germ formation in vivo.
Whole exome sequencing identifies variants causing different monogenic diseases in one nuclear family. Y. Li1, E. Lausch2, K.O. Schwab2, N. van der Werf-Grohmann2, T. Veiten2, D. Lütjohann1, P.V. Lotti1, U. Matysiak-Scholze1, B. Zabel2, A. Köttgen1,4, 1) Department of Internal Medicine IV, University Hospital Freiburg, 79110 Freiburg, Germany; 2) Paediatric Genetics Division, Centre for Paediatrics and Adolescent Medicine, University Hospital Freiburg, 79106 Freiburg, Germany; 3) Institute of Clinical Chemistry and Clinical Pharmacology, University Clinics of Bonn, 53127 Bonn, Germany; 4) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Purpose: Whole exome sequencing has greatly facilitated the identification of mutations causing single-gene disorders, and has the potential to implicate genes previously not known to cause monogenic diseases. The purpose of this study was to search for causal mutations in two children from a consanguineous marriage. Both children are affected by hypomagnesemia and congenital hypothyroidism; one sibling also displayed hyperlipidemia. Both children are treated with thyroxine since shortly after birth and elevated serum plant sterols. Mutations in both TRPM6 and ABCG5 were confirmed by Sanger sequencing. We are currently following up additional variants that fulfilled the filtering criteria as candidates for the thyroid phenotype.

Results: Across individuals, mean coverage of the target region ranged from 40x to 63x, and target coverage at >20x ranged from 83% to 91%. After variant filtering (<1% population frequency; stop, splice, frameshift or missense variants; autosomal recessive mode of inheritance; and region of homozygosity obtained homozygosity mapping), 19 candidate variants were identified. One of these is the c.2667+1G>A mutation in TRPM6, which encodes a magnesium transporter expressed in intestine and kidney. This splice variant leads to skipping of exon 19 and is known as autosomal recessive hypomagnesemia with secondary hypocalcemia (MIM #602014). In the child with hyperlipidemia, we identified a homozygous c.1336T>C mutation in ABCG5, a sterol transporter. This known mutation introduces a premature stop codon and causes autosomal recessive sitosterolemia (MIM #210250). The other family members are heterozygous for this mutation.

Biochemical analyses confirmed the diagnosis of sitosterolemia with elevated serum plant sterols. Both children are treated with thyroxine since shortly after birth and elevated serum plant sterols. Mutations in both TRPM6 and ABCG5 were confirmed by Sanger sequencing. We are currently following up additional variants that fulfilled the filtering criteria as candidates for the thyroid phenotype. Both children are treated with thyroxine since shortly after birth and develop normal thyroid function. Differences in lipidlowering treatment can be adapted for the specific diagnosis.
Mutations in PIK3R1 cause SHORT syndrome. D. Dyment1, A. Smith2, D. Alcantara3, J.A. Schwartzentruber4, L. Basel-Vanagaite5, C.J. Curry6, I.K. Temple7, W. Readon8, S. Mansouri9, M.R. Haq9, R. Gilbert9, O.J. Lehmann11, M.R. Vanstone2, C.L. Beaulieu2, Th. FORGE CANDADA CONSORTIUM12, J. Mejajek2, D.E. Bulman2, M. O’Driscoll2, K.M. Boycott13, A.M. Innes10,14. 1) Children’s Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Children’s Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada; 3) Genome Damage and Stability Centre, University of Sussex, Brighton, UK; 4) McGill University and Genome Quebec Innovation Centre, Montréal, Quebec, Canada; 5) Department of Pediatric Genetics, Schneider Children’s Medical Center of Israel, Petah-Tikva, Israel; 6) Genetic Medicine Central California, Fresno, CA, USA and the Department of Pediatrics, University of California San Francisco, CA, USA; 7) 7Faculty of Medicine, University of Southampton and the Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Southampton, UK; 8) Our lady’s Hospital for Sick Children, Crumlin, Dublin 12, Ireland; 9) SW Thames Regional Genetics Service, St. George’s Hospital Medical School, London, UK; 10) Department of Paediatric Nephrology, Southampton Children’s Hospital, University Hospital Southampton University Hospitals NHS Foundation Trust, Southampton, UK; 11) Department of Ophthalmology, University of Alberta, Edmonton, AB, Canada; 12) Steering Committee Membership is listed at end of poster; 13) Department of Medical Genetics, University of Calgary, Calgary, AB, Canada; 14) Alberta Children’s Hospital Research Institute for Child and Maternal Health, University of Calgary, Calgary, AB, Canada.

SHORT syndrome is a rare, multisystem disease characterized by short stature, anterior chamber eye anomalies, characteristic facial features, lipodystrophy, hernias, hyperextensibility and delayed dentition. As part of the FORGE Canada Consortium we studied individuals with clinical features of SHORT syndrome to identify the genetic etiology of this rare disease. Whole exome sequencing in an affected child-unaffected parents triad identified a de novo truncating mutation, c.1971T>G, p.Tyr657*, in another affected individual.

We also report the phenotypes in an unrelated family. The other mutation was a de novo truncating mutation, p.Asn636Thrfs*18. Heterozygous mutations in exon 14 of PIK3R1 were subsequently identified by Sanger sequencing in three additional affected individuals and was also identified and shown to segregate with the phenotype in an unrelated family. The other mutation was a de novo truncation mutation, c.1971T>G, p.Tyr657*, in another affected individual. PIK3R1 is involved in the PI-3K-AKT-mTOR pathway. Our findings show that PIK3R1 mutations are the major cause of SHORT syndrome and suggest the molecular mechanism of disease may involve down-regulation of the PI-3K-AKT-mTOR pathway.
Protein 3, is associated with a rare subtype of frontonasal dysplasia. A mutation in ATLE3 (NM_005078:exon15:c.A1517G:p.K506R), encoding the transducin-like enhancer protein 3, is associated with a rare subtype of frontonasal dysplasia. Y. Xi, C.L. Beaulieu, A.C. Smith, J. Schwartzentruber, F.P. Favaro, D.B. Bulman, M.A. Guion-Almeida, A. Richieri-Costa, A.M. Innes, K.M. Boycott FORGE Canada Consortium. 1) Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montréal, Quebec, Canada; 3) Department of Clinical Genetics, Hospital of Rehabilitation of Craniofacial Anomalies, University of Sao Paulo (HRAC-USP), Bauru, SP, Brazil; 4) Department of Medical Genetics, University of Calgary, Calgary, AB, Canada.

Frontonasal dysplasia is the hallmark of several syndromes involving the frontonasal process. One of the syndromes within this spectrum is characterized by midline facial clefting, agenesis of the corpus callosum, basal encephalocele, pituitary and ocular abnormalities. These features are very close to Sakoda syndrome (OMIM # 610871). All cases reported to date have been sporadic. Here employing whole-exome sequencing on a patient and his parents, we identified a de novo mutation in the TLE3 gene (NM_005078:exon15:c.A1517G:p.K506R), encoding the transducin-like enhancer protein 3. Analysis of a further three patients with a similar clinical presentation did not identify any mutations in this gene and suggests this may be a genetically heterogeneous developmental syndrome. TLE3 belongs to the Grouth and Development Enamel family, members in which are broadly expressed during development and have essential functions in many developmental pathways including Notch and Wnt signalling (Jennings et al Genome Biol 2008).

TLE family proteins perform their transcriptional repression functions by interacting with a diverse profile of transcription factors (including Hes, Runx, and Pax) rather than binding DNA directly. The C-terminal WD repeat domain in TLE proteins is highly conserved and responsible for their interactions with transcription factors (Carvalho et al Mol Endocrinol 2010). In our study, the mutation in TLE3 (p.K506R) lies in the first WD repeat and this mutation could change its binding efficiency to various transcription factors during frontonasal development. Recently, TLE proteins were also shown to biochemically interact with EphrinB1, mutations in which cause craniofrontonasal dysplasia; the WD domain plays important role in regulation of this interaction (Kamata et al BMB Reports 2011). Thus TLE3 is a novel candidate gene for this subtype of frontonasal dysplasia. Future experiments such as immunoprecipitation and Western blotting will be performed to dissect the binding capacities of wt and mutant TLE3 with different transcription factors and EphrinB1. Morpholino knockdown on zebrafish tle3a and tle3b (TLE3 homologs in zebrafish) will also be performed to characterize their functions during craniofrontonasal development.
3075T
Exome Sequencing Identifies Germline Mutations in SPAG1 as a Cause of Primary Ciliary Dyskinesia Associated with Defective Outer and Inner Dynein Arms. M.A. Zarivach1, L.E. Ostrowski2, N.T. Loges3, T. Hurd4,5, M.W. Leigh6, L. Huang7, W.E. Wolf8, J.L. Carson9, M.J. Hazuka9, W. Yin10, S.D. Davis11, D.S. Dell12, T.W. Ferkol12, S.D. Sage13, K.N. Olivier14, C. Jahnke15, H. Olbrich16, C. Werner17, H.Y. Gee18, E.A. Otto19, J. Halbritter4,20, E.H. Turner11, A.P. Lewis13, M.J. Bamshad13,14, D.A. Nickerson13, F. Hildebrandt11,12,16, J. Shendure15, H. Omran2, M.R. Knowles2, Genetic Disorders of Mucociliary Clearance Consortium (GDMCC). 1) Department of Pathology & Laboratory Medicine, UNC School of Medicine, Chapel Hill, NC 27599, USA; 2) Department of Medicine, UNC School of Medicine, Chapel Hill, NC 27599, USA; 3) Department of General Pediatrics and Adolescent Medicine, University Hospital Muenster, 48149 Muenster, Germany; 4) Department of Pediatrics, University of Michigan, Ann Arbor, Michigan 48109, USA; 5) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK; 6) Department of Pediatrics, UNC School of Medicine, Chapel Hill, NC 27599, USA; 7) Section of Pediatric Pulmonology, Allergy and Sleep Medicine, James Whitcomb Ruket Riley Hospital for Children, Indiana University School of Medicine, Indianapolis, IN 46202, USA; 8) Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON, M5G1X8, Canada; 9) Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA; 10) Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO 80045, USA; 11) Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA; 12) Division of Nephrology, Harvard Medical School, Boston Children’s Hospital, Boston, MA 02114, USA; 13) Department of Human Genetics, University of Washington School of Medicine, Seattle, WA 98195, USA; 14) Department of Pediatrics, University of Washington School of Medicine, Seattle, WA 98195, USA; 15) Howard Hughes Medical Institute, Chevy Chase MD 20815, USA.

Primary ciliary dyskinesia (PCD) is a genetically heterogeneous, autosomal recessive disorder, characterized by oto-sonto-pulmonary disease and situs abnormalities. PCD-causing mutations have been identified in 19 genes, but collectively they account for only ~65% of all PCD. To identify additional genes that cause PCD, we performed exome sequencing on 3 unrelated probands with ciliary outer and inner dynein arm (ODA+IDA) defects. Mutations in SPAG1 were identified in one family with three affected siblings. Further screening of SPAG1 in 98 unrelated affected individuals (62 with ODA+IDA defects; 35 with ODA defects: one without available ciliary ultrastructure) revealed biallelic loss-of-function mutations in 11 additional individuals (including one sib-pair). All 14 affected individuals with SPAG1 mutations had a characteristic PCD phenotype, including 9 with situs abnormalities. Additionally, all individuals with mutations who had defined ciliary ultrastructure had ODA+IDA defects. SPAG1 was present in human airway epithelial cell lysates, but was not present with isolated axonemes, indicating that it likely plays a role in the cytoplasmic assembly/trafficking of the axonemal dynein arms. Together, these results demonstrate that mutations in SPAG1 cause PCD with ciliary ODA+IDA defects, and that exome sequencing is useful to identify genetic causes of heterogeneous recessive disorders. This abstract was funded by 5U54HL096458-06 (NIH/ORDR/NHLBI), 5R01HL071796 (NIH/NHLBI), 5R01HL084976 (NIH/NHLBI), RC2 HL-102923 (NIH/NHLBI), S21HG004749 (NIH/NHGRI), U1L TR000083 (NIH/NCATS), U1L TR000154 (NIH/NCATS), DFG Om2/009/12, BESTCILIA (EU), SYSCILIA (EU) and Howard Hughes Medical Institute.

3076F
A Genetic Snapshot of Risk Alleles in the Jewish Population. D. Zielinski, Y. Erlich. Whitehead Institute for Biomedical Research, Cambridge, MA. Carrier screening has been instrumental in preventing rare genetic disorders. As a result of various bottlenecks and isolation, risk alleles for certain disorders are found at a higher frequency in the Jewish population, which has particular benefits from prenatal screening. Despite great success, including the prevention of fatal disorders in some communities, carrier screening does not cover the entire allele spectrum, resulting in false negatives and thus residual risk. Current methods to ascertain new risk alleles rely on sequencing affected children. We are taking a proactive approach to uncover risk alleles through complete sequencing of the entire coding regions of more than 150 genes documented by the Israeli Ministry of Health that are associated with genetic disorders in Jews of various ethnic backgrounds. The aims of this study are to uncover novel alleles in known disease-associated genes and generate more data on disease burden, including carrier frequencies of known founder mutations as well as minor mutations.

Studying rare variations is a needle in a haystack problem, as large cohorts have to be sequenced in order to trap the variations and gain statistical power. However, sample preparation techniques have not kept pace with the exponential growth in high throughput sequencing technology and rely on tedious, individual barcoding. We have developed a scalable solution, called DNA Sudoku, to reduce sample preparation time and costs and increase throughput. After several successful pilot studies, we have employed the pipeline to take a genetic snapshot of risk alleles for genetic disorders in the Jewish population, focusing our efforts on a relatively small cohort of healthy, ethnically matched individuals, in order to create a comprehensive risk catalog and further improve carrier screening.
3077W Exome sequencing identifies de novo and post-zygotic mutations in GATA4 associated with congenital diaphragmatic hernia. J. Bennett1, L. Yu2, Y. J. Jung3, Y. C. Chu4, J. P. Rozycki5, T. Cronin6, K.S. Azarow7, F.Y. Lim8, D.H. Chung9, D. Potoka10, B.W. Warner10, B. Bucher10, C. Stolar10, G. Aspelin10, M.S. Arkovitz11, H. Mofford11, W.K. Chung11, The University of Washington Center for Mendelian Genomics. 1) Department of Medical Genetics, University of Washington, Seattle, WA; 2) Division of Molecular Genetics, Department of Pediatrics, Columbia University Medical Center, New York, NY; 3) Department of Biomedical Informatics, Columbia University Medical Center, New York, NY; 4) Department of Pediatrics, University of Michigan, Ann Arbor, MI; 5) Colorado Fetal Care Center, Division of Pediatric General, Thoracic, and Fetal Surgery, Children’s Hospital Colorado and the University of Colorado School of Medicine, Aurora, CO; 6) Department of Pediatric Surgery, University of Nebraska College of Medicine, Omaha, NE; 7) Division of Pediatric General, Thoracic, and Fetal Surgery, Center for Molecular Fetal Therapy, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH; 8) Department of Pediatric Surgery, Vanderbilt University Medical Center, Nashville, TN; 9) Department of Pediatric Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA; 10) Division of Pediatric Surgery, Washington University School of Medicine, St. Louis, MO; 11) Division of Pediatric Surgery, Department of Surgery, Columbia University College of Physicians and Surgeons, New York, NY; 12) Division of Pediatric Surgery, Tel Hashomer Medical Center, Tel Hashomer, Israel.

Congenital diaphragmatic hernia (CDH) is a relatively common and severe birth defect that affects 1 in 3,000 births. It is characterized by herniation of abdominal viscera into the chest cavity through the incomplete formation of the diaphragm. It is a frequent cause of neonatal death and contributes significantly to its morbidity and mortality. Although chromosomal anomalies and mutations in several genes have been implicated in the etiology of CDH, the cause for most patients is unknown. We used whole exome sequencing (WES) in two families with CDH and congenital heart disease, and identified mutations in GATA4 in both. In the first family, we identified a de novo missense mutation (c.1386C>T; p.R456C) in a sporadic CDH patient with tetralogy of Fallot. In the second, a nonsense mutation (c.715G>T; p.V239X) was identified in two siblings with CDH, with veno-ventriculoseptal defect. The G238X mutation was inherited from their mother, who was clinically affected with congenital absence of the pericardium, patent ductus arteriosus, and gastrointestinal malrotation, but without clinical manifestations of CDH. Deep sequencing of the GATA4 gene in two siblings with GDAP1 associated with PHARC syndrome, revealed significant homozygous GDAP1 variants. Moreover, Sanger sequencing of the breakpoint revealed this deletion to be identical to one reported previously by Fiskerstrand et al. (2010) in a family from the United Arab Emirates. This finding suggests that this deletion might have occurred as a single event in a common ancestor's haplotype and that it might currently be segregating in the population of this region. The deletion breakpoint maps within a region of microhomology and high identity between a pair of flanking Alu elements. Interestingly, WES data can reveal copy number variation of direct clinical relevance, providing additional utility of exome sequencing in identifying the genetic basis of rare genetic disorders.

3079F Depletion of exome sequencing reads reveals a 14kb homozygous deletion in a patient with PHARC syndrome. T. Harel1, C. Gonzalez-Laustri2, T. Cranved3, M. Koenig3, R.A. Gibbs4,5, J.R. Lupski6,7,8, J.B. Hopkins Center for Mendelian Genomics. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; 2) Departments of Neurology and Pediatrics, Johns Hopkins University, Baltimore, Maryland; 3) Institut de Genetique et de Biologie Moleculaire et Cellulaire (IGBMC), CNRS-INSERM-Universite de Strasbourg, 67404 Illkirch, France; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, 77030; 6) Texas Children’s Hospital, Houston, TX, 77030.

Whole-exome sequencing (WES) has proved extremely useful in identifying simple nucleotide variation (SNV), i.e., single base pair mutations and small indels, responsible for highly penetrant Mendelian phenotypes. However, assessment of copy number variation (CNV), small deletions and duplications, responsible or contributing to Mendelian phenotypes from WES data is still a task in progress as bioinformatic algorithms are being developed and improved. We performed exome sequencing in a patient from a Qatari consanguineous family that presented with polyneuropathy, hearing loss, ataxia, retinitis pigmentosa, and cataract (PHARC). Analysis of the WES data reads and mapping in the vicinity of the ABHD12 gene, encoding abhydrolase domain-containing protein 12 and associated with PHARC syndrome, showed complete absence of sequencing reads in a 14kb region on chromosome 20, encompassing exon 1 of ABHD12. Additionally, there was absence of heterozygosity (AOH) surrounding the proposed deletion region. Subsequent confirmation by PCR of the deletion breakpoint confirmed autosomal recessive segregation of the deletion with the phenotype in the family. Moreover, Sanger sequencing of the breakpoint revealed this deletion to be identical to one reported previously by Fiskerstrand et al. (2010) in a family from the United Arab Emirates. This finding suggests that this deletion might have occurred as a single event in a common ancestor’s haplotype and that it might currently be segregating in the population of this region. The deletion breakpoint maps within a region of microhomology and high identity between a pair of flanking Alu elements. Interestingly, WES data can reveal copy number variation of direct clinical relevance, providing additional utility of exome sequencing in identifying the genetic basis of rare genetic disorders.

3090W From monogenic to oligogenic: strategies to uncover oligogenic modes of inheritance in individuals with intellectual disability. W.D. Jones1, M. Van Kogelenberg2, D. King1, T. Fitzgerald1, H.V. Firth1,2, M.E. Hurles1, J.C. Barrett2, Deciphering Developmental Disorders Study. 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Cambridge University Department of Medical Genetics, Addenbrooke’s Hospital, Cambridge CB2 0QQ.

Phenotypes that segregate in a Mendelian manner have long been thought to result from alterations in a single gene or unique locus. However, there is recent evidence to suggest that oligogenic modes of inheritance may underlie intellectual disability in some individuals. The Deciphering Developmental Disorders (DDD) Project is a study of individuals with developmental disorders with recruitment from 24 Regional Genetics Services Networks in the United Kingdom and Republic of Ireland. Individuals first undergo detailed clinical phenotyping at their local Genetics centre. This is followed by exome-array Comparative Genomic Hybridisation (CGH) of the proband and exome sequencing of the proband and both parents. The combining of data from array CGH analysis and WES sequencing allows a powerful platform to search for evidence of oligogenic modes of inheritance in individuals with intellectual disability. A control data set known as the ‘imaginary siblings’ will be generated to carry out this analysis, consisting of variants present in the parents not inherited by the proband. The computational and statistical methods employed in the analysis for oligogenic models will be presented alongside preliminary results.
3081T

Integrative genomics reveals that distal cis-regulatory mutations cause isolated pancreatic agenesis. M.N. Weedon,1, I. Cebolja2-3, A. Patch,1, S.E. Flanagan1, E. De Franco,1 R. Caswell1, S.A. Rodriguez-Segui3,4, C. Shaw-Smith1, C. Cho5, H. Lango Allen1, J.A.L Houghton1, C.L. Roth4, R. Chen1, K. Hussain3,4,5, P. Marsh10, L. Vallier9, A. Murray1, S. Ellard1, J. Ferrer2-3,11, A.T. Hattersley1. 1) University of Exeter Medical School, UK; 2) Institut d’Investigacions Biomèdiques August Pi i Sunyer, Spain; 3) CIBER de Diabetes y Enfermedades Metabólicas, Spain; 4) Ciudad Universitaria, Argentina; 5) Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute, Cambridge, UK; 6) Seattle Children’s Hospital Research Institute, USA; 7) School of Biomedical Science, King’s College London, UK; 8) London Centre for Paediatric Endocrinology and Metabolism, London, UK; 9) University College London, London, UK; 10) School of Medicine, King’s College London, London, UK; 11) Imperial College, London, UK.

The contribution of cis-regulatory mutations to human disease remains poorly understood. Whole genome sequencing can identify all non-coding variants, yet discrimination of causal regulatory mutations represents a formidable challenge.

We performed homozygosity mapping and whole genome sequencing on probands from two consanguineous families with non-syndromic pancreatic agenesis. We first looked for recessive coding mutations in the exomes of these patients, but no causal mutations were identified.

To search for non-coding disease-causing mutations, we annotated the 6,024 rare (<1% frequency in the 1000 genomes project) or novel homozygous variants from these patients using epigenome maps from human pancreatic progenitor and control cells. Only one variant, a novel SNV, occurred in a functionally annotated region and was shared by both unrelated patients. This variant occurs several tens of kb from PTF1A in a short (~400bp) region of conservation. Follow-up sequencing identified four different recessive point mutations and an 8kb deletion spanning the functional element in 9 of 12 non-syndromic pancreatic agenesis cases.

Experiments including 3C and reporter assays clearly demonstrate that this previously uncharacterised non-coding element acts as a developmental enhancer of PTF1A in human pancreatic progenitor cells. The 6 mutations prevent enhancer activity by abolishing transcription factor binding or deleting the enhancer sequence.

Integrating genome sequencing and epigenomics in a disease-relevant cell type can uncover novel non-coding elements underlying human development and disease.

3082F

Clinical Exome Sequencing identifies de novo mutations in the MLL gene causing Atypical Wiedemann-Steiner Syndrome in two unrelated individuals. F. Quintana-rivera1, S.P. Strom1, J. Mann1, R. Lozano2, H. Lee3, N. Dorrani1, O. O’Lague4, N. Mans5, J.I. Deignan1, E. Vilain4,2,3, S.F. Nelson1,2,3, W.W. Grody2,3,5, 1) Department of Pathology and Laboratory Medicine, UCLA School of Medicine, Los Angeles, CA; 2) Department of Human Genetics, UCLA School of Medicine, Los Angeles, CA; 3) Department of Pediatrics, UCLA School of Medicine, Los Angeles, CA; 4) Kaiser Permanente, Fresno Medical Center; 5) Department of Pediatrics, University Of California Davis Medical Center, Sacramento, CA.

Wiedemann-Steiner Syndrome (WSS) is characterized by mild to moderate developmental delay, dysmorphic facial features, and hypertrichosis cubiti (excessive hair on the elbows) [MIM#: 605130]. Here we report two unrelated patients for whom Exome sequencing of parent-proband trios was performed clinically at the UCLA Molecular Diagnostics Laboratories. For patient 1, clinical features at 9 years of age, include developmental delay, craniofacial and multiple minor anomalies: down-sloping palpebral fissures, thick eyebrows and hair, premature dental eruption, fifth finger clinodactyly, tapered fingers, and astigmatisum. Patient 2 presented at 1 year of age with intrauterine growth retardation, developmental delay, significant central hypotonia, RT. microphthalmia and small palpebral fissure, prominent nasal bridge, micrognathia, Lt. hand 3-4 syndactyly, mild pectus excavatum, and small sacral hair tuft. In patient 1, a novel missense (p.Cys1448Arg) variant was identified in the MLL gene [RefSeq transcript ID: NM_001197104.1]. In patient 2, a de novo splice site mutation (c.4086+1G>A) was identified in the same gene. MLL - Myeloid/Lymphoid Or Mixed Lineage Leukemia Gene - encodes a broadly expressed DNA-binding protein that mediates gene expression in response to cytokines and is located on chromosome 11q23. The region is enriched for target genes, including genes in the HOX and WNT pathways. It is frequently translocated in hematological malignancies, including ALL, AML, and MLL. Mice carrying a hemizygous knockout allele of MLL are small at birth and have reduced growth, indicating that dosage of the gene is critical to development. However, the molecular mechanism of MLL-related WSS has not been established. Based on clinical and molecular genomic findings, both patients may have novel presentations of WSS. As the hallmark hypotonia was not initially appreciated in either case, this syndrome was not suspected clinically. This report expands the phenotypic spectrum of clinical phenotypes and de novo mutations of the MLL gene associated with WSS.

3083W

Follow-up of diagnostic exome sequencing in persons with severe intellectual disability; Re-analysis of data, recurrence screening, CNV detection and genome sequencing. L. Vissers1, H. IJtena1, J. Hehir-Kwa1, C. Gillissen1, J. de Ligt1, R. Leach1, R. Tarar2, W. Nillesen1, B. van Bon1, M. Willemsen1, H. Scheffer1, H. Brunner2, B. de Vries1, T. Kleefstra3, J. Veelmann1. 1) Dept Human Gen 855, UMC Nijmegen, Nijmegen, Netherlands; 2) Complete Genomics, Inc. 2071 Sterlin Court Mountain View, CA 94043.

In 2012 we described the use of exome sequencing in the diagnostic workup of patients with unexplained severe intellectual disability (ID). In this initial study we detected 16 pathogenic mutations in 100 patients, mostly involving de novo germline point mutations and insertions events. Here we describe our ongoing genetic studies in this cohort. One of the follow-up studies is aimed at identifying recurrent de novo germline mutations in 21 candidate genes by using various approaches, including targeted next generations sequencing in a large ID cohort. So far these efforts have revealed additional mutations in 7 of our candidate ID genes. Detailed clinical analysis of the patients with mutations showed phenotypic overlap, confirming that these genes cause ID when mutated. A second follow-up study involves the re-analysis of the existing exome data of these patients and their parents using a new software program, Lifescope, which is more sensitive in detecting variants than the program originally used. This analysis provided a positive molecular diagnosis in 5 patients by the identification of de novo mutations in known ID genes. A third follow-up study, which is currently ongoing, is aimed at the identification of CNVs based on read-depth information of exome data using CoNIFER, a software tool that we recently validated in a series of patients with pathogenic CNVs. Finally, we performed whole genome sequencing in 50 of the unexplained ID patients and their parents. Preliminary results of 2 de trios confirm the presence of novel exonic mutations detected only by re-analysis of the original exome sequencing data. In addition whole genome sequencing in these trios revealed 5 de novo exonic mutations in novel candidate ID genes as well as one de novo mutation in the UTR of a known ID gene. While much work is still ongoing, it does already demonstrate that the initial diagnostic yield of 16% by exome sequencing is an underestimation; at this moment the number is already at 26%. This yield can be substantially higher by further improving variant detection on exome sequencing data by follow-up of the remaining candidate genes in larger clinical cohorts. In addition, whole genome sequencing may not only give a more complete picture of variation present in the exome but will also provide us insight into possible causative mutations in the non-coding part of the genome.

Introduction Adams-Oliver syndrome (AOS) is a rare genetic disorder characterized by aplasia cutis congenita (ACC) and terminal transverse limb defects. Mutations in the RBPJ and ARHGAP31 genes have been described as a cause of autosomal dominant AOS, while mutations in the DOCK6 and EOGT genes were shown to be responsible for the autosomal recessive form. We performed mutation analysis of the RBPJ gene in a cohort of AOS patients to determine the relative contribution of RBPJ to AOS. Materials and Methods A cohort of 58 patients with clinical characteristics of AOS were included in the study, including familial and sporadic cases. Mutation analysis of the RBPJ gene was performed by sequence analysis of all coding exons. Results and discussion In three families an RBPJ missense variant was detected, including one previously reported pathogenic RBPJ mutation (p.Thr441Thr, p.Lys169Glu, p.Phe66Val). Three other intronic variants need further evaluation in a larger control population. In conclusion, mutations in the RBPJ gene seem only present in a limited number of AOS patients.

3085F Application of targeted next-generation sequencing in the diagnosis of pediatric neurological disorders. N. Okamoto1, M. Fuya2, T. Tsunoda2, M. Kato3, S. Saitoh4, M. Yamaski2, Y. Kanemura2,5, K. Kosaki6, 1) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 2) Laboratory for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 3) Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan; 4) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 5) Department of Pediatric Neurosurgery, Takatsuki General Hospital, Osaka, Japan; 6) Division of Regenerative Medicine, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 7) Department of Neurosurgery, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 8) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

Introduction We have developed a next generation sequencing (NGS)-based mutation screening strategy. This system enables us to screen 284 genes associated with pediatric neurological disorders. [Materials and Methods] Fifty patients with neurological disorders were included in the study. They are from one institution (Osaka Medical Center). Under the approval by our institutional ethics committee, patients were analyzed by the targeted exome sequence capture. To capture the target exonic DNAs, we used SureSelectXT Custom capture library (Agilent) for 1.6 Mb exons of neuronal genes. The sequence library was constructed using SureSelect XT Target Enrichment System for Illumina Paired-End Sequencing Library kit (Agilent). We performed sequencing the DNA with 101-bp paired-end by 24-multiplex per lane using Illumina HiSeq 2000 sequencer and obtained 1.90 ± 0.123 (mean ± s.d.) Gb sequence data. To extract tissue disease causal mutation, we excluded known variants found in dbsNP, 1000 Genomes Project, ESP6500 and our control samples, and narrowed the candidates to nonsynonymous and nonsense SNVs. [Results] Several pathogenic mutations were identified. An early termination codon in the DRYRK1A gene was found in a patient with severe intellectual disability (ID), absent speech, motor disturbance, and visual disturbance (Patient 1). Heterozygous mutations in the CASK gene were found in two females with mental retardation and microcephaly with pontine and cerebellar hypoplasia (Patient 2,3). A patient with microcephaly and severe ID and showing similar conditions with Kleefstra syndrome had a mutation in the MBDS gene (Patient 4). A patient with cutis laxa had an ALDH18A1 mutation (Patient 5). A female patient with Angelman syndrome like features had mutation in the GABRD gene (Patient 6). A patient with ID, piosis and pachygyria of the frontal lobe had a mutation in the ACTB gene. He was compatible with Baraitser-Winter syndrome (patient7). [Discussion] Our NGS-based mutation screening strategy is sensitive and specific in detecting sequence variants in some neurological disorders. We propose that this NGS-based targeted sequencing method would be an alternative and to current technologies for identifying the multiple genetic causes of neurological disorders.

3086W Disruption of the Rac GTPase activator DOCK7 in epileptic encephalopathy. F.F. Hamdan1, J.M. Capo-Chichi4, B. Maranda2, G.A. Rouleau2, J.M. McAllister3, 1) Centre Hospitalier Universitaire de Sherbrooke, Quebec, Canada; 2) Division of Genetics, Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Quebec, Canada; 3) Montreal Neurological Institute, McGill University, Montreal, QC, Canada.

The genetic basis of epileptic encephalopathy remains unexplained in the majority of cases. Here we used exome sequencing to study two sisters (aged 7 and 9 yrs) with severe intellectual disability and intractable epilepsy from a non-consanguineous French Canadian family. We found 186 rare amino acid changing or splicing variants that were shared between the exomes of the 2 affected siblings. None of these was homozygous. DOCK7 is the only gene containing compound heterozygous mutations (c.3730G>T:p.R1273X and c.2510delA:p.D837fs; NM_001271999) shared by the sisters. Both mutations, which truncate DOCK7 upstream of its GEF domain, were absent from all public SNP databases (1000 Genomes, Exome Variant Server (EVS), dbSNP138), and from our in-house exome database (>1000) which includes at least 200 French Canadian individuals. Inspection of DOCK7 for truncating or splicing variants in the exome variant server data set, revealed only 2 heterozygous truncating variants, in 2 different individuals, out of approximately 13000 sequenced alleles. This indicates that, despite its large size (2129 amino acid), DOCK7 does not accumulate truncating mutations and, therefore, maybe essential in humans. DOCK7 encodes a guanine exchange factor (GEF) belonging to the DOCK180 super family whose members activate Rho GTPases. DOCK7 is enriched in the brain and has been shown to activate Rac and promote axon formation, axon myelination, and cortical neurogenesis. Based on the importance of DOCK7 in brain function and the severity and rarity of the identified mutations, our data suggests that disruption of DOCK7 may be responsible for the epileptic encephalopathy observed in this family. Interestingly, DOCK7 is known to physically interact with the TSC1-TSC2 protein complex whose dysfunction causes Tuberous Sclerosis Complex, a multisystemic disorder with high incidence of ID and epilepsy. Identification of additional similarly affected families with DOCK7 recessive deleterious mutations will be necessary to establish DOCK7 as a neurodevelopmental disease gene. Collaborations towards this effort are welcome.

3087T Whole genome sequencing, clinical interpretation, and deep brain stimulation in a severely mentally ill person. S. O’Brien1, S.S. Vinette1, S. Telang1, J.J. Jacobson2, J.L. Lowrey3, E. Kiruluta1, J.M. Capo-Chichi4, G. Higgins5, M. Reese3, 1) Stanford Institute for Genetic Cognitomics, One Buntington Road, Cold Spring Harbor Laboratory, NY, USA, 11724; 2) Stony Brook University, 100 Nicolls Rd, Stony Brook, NY, USA, 11794; 3) Utah Foundation for Biomedical Research, E 3300 S, Salt Lake City, Salt Lake City, UT, USA, 84106; 4) Omicia Inc., 2200 Powell St., Emeryville, CA, USA, 94608; 5) AssureRx Health, Inc., 6030 S. Mason Montgomery Road, Mason, Ohio 45040.

We report here the detailed phenotypic characterization, clinical-grade whole genome sequencing (WGS), and two year outcome of one man with severe obsessive compulsive disorder treated with deep brain stimulation (DBS) targeting the anterior limb of the internal capsule (ALIC). Since implantation, this man has reported steady improvement, highlighted by a drop in his Yale-Brown Obsessive Compulsive Scale score from ~35 to a score of ~25. A rechargeable Activa RC neurostimulator battery has been of major benefit in terms of facilitating a degree of stability and control over the stimulation. His psychiatric symptoms reliably worsened within hours of the battery becoming depleted, thus providing concomitant evidence for the efficacy of DBS for OCD in this person. Whole genome sequencing in the CLIA-certified Illuma WGS lab revealed that he is a heterozygote for the p.Val166Met variant in BDNF, encoding a protein that is a member of the neurotrophic factor family, and which has been found to predispose carriers to various psychiatric illnesses. He carries the p.Glu429Ala allele in methylentetrahydrofolate reductase (MTHFR) and the p.Asp7Asn allele in ChAT, encoding choline O-acetyltransferase, which synthesizes the neurotransmitter acetylcholine, with both alleles having been shown to confer an elevated suicide risk in cases of schizophrenia. Whole genome sequencing revealed variants in his genome, including pharmacogenetic variants, and have archived and observed the clinical sequencing data to him, so that he and others can re-analyze his genome for years to come. To our knowledge, this is the first N=1 human study of the clinical neuroscience of WGS with management of genetic results for a person with severe mental illness and 2) detailed neuropsychiatric phenotyping and individualized treatment with deep brain stimulation for his OCD. His WGS results and positive outcome with DBS for OCD is one example of how genomic medicine, including genomics-guided preventive efforts and brain-implantable devices.
3088F
Genes Make Sense: Seeking Causative Genes for Human Congenital General Anosmia. A. Alkelai1, T. Olender1, P. Tatarsky2, V. Boyko1, D. Oz-Levi1, I. Keydar1, R. Milgrom1, E. Feldmesser1, E. Ben-Asher1, E. K. Ruzzo2, D. B. Goldstein2, E. Pras3,4, D. Lancet1. 1) Molecular Genetics, The Weizmann Institute of Science, Rehovot, Israel; 2) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708, USA; 3) The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Ramat Gan, Israel; 4) The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

The neuronal and cell-biological mechanisms underlying vertebrate olfaction have been studied in considerable detail, but there is a gap in our understanding of the genetic and molecular basis of chemosensory variability and related monogenic disorders. Congenital general anosmia (CGA) affects ~0.1% of general population, and appears in isolated or syndromic forms. While for an appreciable number of syndromic CGA types a causative gene has been identified, knowledge about the genetic basis of the isolated CGA instances has been scarce. Our laboratory has samples from a large CGA cohort, including 66 families of Jewish origin (Feldmesser et al, Chem Senses 32:21-30, 2007). We performed whole-exome capture (SureSelect Human All Exon kit - Agilent Technologies, Santa Clara, CA) and next-generation sequencing (Hiseq 2000 sequencing system, Illumina, Inc, San Diego, CA) in 22 selected individuals from 7 of these families multiply affected with anosmia. Now we report results of the Affymetrix human SNP 6.0 array genotyping, linkage analysis, homozygosity mapping, CNV analysis and whole-exome sequencing in one of these families with X-linked/recurrent/ dominant with incomplete penetrance mode of inheritance. By using variety of methods we identified number of candidate regions harboring possible pathogenic variants. The best candidate variants are located in such genes as SPATA5L1 (compound heterozygous mutation), CALML6 (heterozygous mutation) and SEMA3A (heterozygous mutation) and now are being further validated. All the implicated genes appear in our recently constructed general olfactory sensitivity database (GOSdb) (Keydar et al, Hum Mutat, 2013). The identification of specific pathogenic functional CGA variants will help to elucidate the molecular basis of general olfactory sensitivity.

3089W
Genetic mapping and whole exome sequencing to unravel the genetic basis of undiagnosed non-syndromic arthrogryposis multiplex congenita. 1) Molecular Genetics, 2) Biogenetics, 3) Center for Developmental Biology and Tissue Engineering, 4) Biocat Research Centre, 5) Center for Genomics and Computational Biology, 6) Biocat Research Centre, 7) Department of Neurology and Neurosurgery, 8) La Timone Hospital, 9) Department of Neurology, 10) La Timone Hospital, 11) Department of Neurology, 12) La Timone Hospital, 13) Department of Neurology, 14) La Timone Hospital, 15) Department of Neurology, 16) La Timone Hospital, 17) Department of Neurology, 18) La Timone Hospital, 19) Department of Neurology, 20) La Timone Hospital, 21) Department of Neurology, 22) La Timone Hospital, 23) Department of Neurology, 24) La Timone Hospital.

Arthrogryposis multiplex congenita (AMC) is characterized by congenital contractures of at least two distinct joints of the body. The overall incidence is 1 in 3000 live births. Non-syndromic AMC is the direct consequence of fetal hypo/akinesia sequence which may lead, in addition to AMC, to pterygia, lung hypoplasia, diaphragmatic defect, or cleft palate. Non-syndromic AMCs include a large spectrum of diseases of motor neurons, neuromuscular junction, or skeletal muscle. The difficulty in establishing the genetic diagnosis in AMC patients is likely caused by the high locus heterogeneity and/or non-identified disease genes and the lack of suitable screening methods. In order to gain insight into the underlying cause of these diseases, whole genome scanning using SNP microarrays alone or combined with whole exome sequencing (WES) were performed in a cohort of 32 multiplex and/or consanguineous families with undiagnosed non-syndromic AMC. In these families, exome and/or whole-genome sequencing lead not led to established diagnosis. We identified mutations in genes already known to be responsible for AMC or neuromuscular disorders (NMD) in 20 out of 32 (62%) families. The primary targets were skeletal muscle in 12 families, neuromuscular junctions in 4 families, axoglial in 3 families or other target in 1 family. Among known AMC or NMD genes, our study confirmed that mutations of SYNE-1 are responsible for AMC and revealed two AMC families carrying mutations of the TTN gene extending the clinical spectrum of TTN gene mutations. The few mutations of 4 new genes were identified. The mutations of these genes may be classified as causative (2 genes), highly candidate (1 gene) or possibly causative (1 gene). This is based on gene mutation type, RNA and morphological analysis of the neuromuscular system (including transmission electron microscopy) and the availability of mouse models or the generation and characterization of a zebrafish model. NGS technologies may be therefore regarded as the most promising approach to unravel undiagnosed AMC leading to a diagnosis in ~80% of cases as shown in the present study. This will be particularly useful during pregnancy when ultrasound examination reveals AMC or reduced fetal mobility but also when single joint contracture is associated with additional symptoms such as the ones described above. Establishing early diagnosis should provide accurate information on the prognosis.
Lack of access to new technologies inhibits discoveries. In an effort to equalize access to next generation sequencing (NGS) for its researchers and clinicians, Boston Children’s Hospital (BCH) awarded a number of pilot grants to provide whole genome and exome sequencing for patients and human research subjects. Emphasis was on assessing the impact of NGS on analysis of a mix of patients with phenotypes ranging from severely affected individuals with diagnostic dilemmas and for whom testing would have a likelihood of clinical impact, to subjects or cohorts of subjects with phenotypes that were likely to yield important new insights into the biology and treatment of disease. Awards were determined by multidisciplinary, cross-departmental committees, and sequencing was performed by the BCH Genetic Diagnostics Laboratory and LabCorp. Datasets were used to develop an in-house BCH NGS analysis pipeline optimized for human disease gene discovery from pedigrees and cohorts. In addition to providing sequencing data to clinicians and investigators, the project provided a single standard of oversight for the hospital, and evaluated the entire process from selecting patients for sequencing, obtaining their informed consent, generating and analyzing sequencing data, interpreting and reporting clinical results, integrating results into the medical record, seeking reimbursement from payors, and complying with governmental regulations. As of May 2013, the pilot has awarded over 600 exomes and genomes for 42 investigators from payors, and complying with governmental regulations. As of May 2013, the pilot has awarded over 600 exomes and genomes for 42 investigators. From the second award, one paper has been submitted, 3 genes have been discovered with one paper published, and two more papers in preparation. In less than 18 months, results of patients now have diagnoses, and two have had treatment changes as the branch point. The proband originally presented with severe muscular atrophy, tongue fasciculations, and ventilator dependency, with no family history of SMA. MLPA testing at another lab had reported a single functional copy of SMN, and gene sequencing was requested at our molecular diagnostic laboratory. Mutations at the branch point of intron 6 have been reported to affect inclusion of exon 7 in the SMN product; however, the position of the insertion precludes detection by the MLPA assay used, as the primers are 3’ to the inserted Alu. Our SMN quantitative assay uses primers within intron 6 and intron 7, and detected zero copies of SMN1, 1 copy of SMN2, and a product of an unusual size that did not appear to digest with Drai. This product was purified, sequenced, and revealed the presence of a normal SMN1 exon 7 preceded by a complete Alu element of the Yb8 family. Members of the Yb8 family of Alu elements are described as currently active in retroposition, and there are reports of de novo Alu Yb8 insertions causing disease. The ‘A’ tail found on the Alu in our patient’s SMN1 gene was 40 bp in length. Recently inserted Alus are characterized by A-tails between 40 and 97bp. Parental testing revealed that the father carries one copy of SMN1, whereas the mother carries two copies of SMN1, one containing the same mutated SMN1 allele, thus confirming the pathogenicity of the Alu insertion in this family. Unique neuroimaging features present in both brothers and not typically seen in 5q SMA include generalized cortical atrophy, hypoplastic appearing optic nerves and possible CNS myelination abnormalities. The severe infantile phenotype in association with congenital weakness and clearly evident central nervous system involvement expand the phenotype previously associated with infantile onset 5q SMA.

3091F
Alu Yb8 insertion near SMN1 exon 7 as a rare cause of SMA. S.M. Kirwin, K.M.B. Vinette, I.L. Gonzalez, S.L. Dugan, K.J. Swoobod, T.M. Newcomb, V.L. Funanage. 1) Molecular Diagnostics Laboratory, Nemo- ours/duPont Hospital for Children, Wilmington, DE; 2) Division of Medical Genetics, Children’s Hospitals and Clinics of Minnesota, Minneapolis, MN; 3) Pediatric Motor Disorders Research Program, Department of Neurology, University of Utah School of Medicine, SLC, UT.

Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disorder primarily affecting young children. Approximately 95% of cases of SMA are caused by homozygous deletion of exon 7 of the SMN1 gene, although up to 5% of cases can arise by heterozygous deletion of a single SMN1 copy paired with a subtle mutation elsewhere within the gene. We report a novel SMN1 gene mutation that involves insertion of an Alu element at an L1 endonuclease recognition site, coinciding with the intron 6 branch point. The proband originally presented with severe muscular atrophy, tongue fasciculations, and ventilator dependency, with no family history of SMA. MLPA testing at another lab had reported a single functional copy of SMN, and gene sequencing was requested at our molecular diagnostic laboratory. Mutations at the branch point of intron 6 have been reported to affect inclusion of exon 7 in the SMN product; however, the position of the insertion precludes detection by the MLPA assay used, as the primers are 3’ to the inserted Alu. Our SMN quantitative assay uses primers within intron 6 and intron 7, and detected zero copies of SMN1, 1 copy of SMN2, and a product of an unusual size that did not appear to digest with Drai. This product was purified, sequenced, and revealed the presence of a normal SMN1 exon 7 preceded by a complete Alu element of the Yb8 family. Members of the Yb8 family of Alu elements are described as currently active in retroposition, and there are reports of de novo Alu Yb8 insertions causing disease. The ‘A’ tail found on the Alu in our patient’s SMN1 gene was 40 bp in length. Recently inserted Alus are characterized by A-tails between 40 and 97bp. Parental testing revealed that the father carries one copy of SMN1, whereas the mother carries two copies of SMN1, one containing the same mutated SMN1 allele, thus confirming the pathogenicity of the Alu insertion in this family. Unique neuroimaging features present in both brothers and not typically seen in 5q SMA include generalized cortical atrophy, hypoplastic appearing optic nerves and possible CNS myelination abnormalities. The severe infantile phenotype in association with congenital weakness and clearly evident central nervous system involvement expand the phenotype previously associated with infantile onset 5q SMA.
Identification of a novel GATA3 mutation in a Taiwanese family with idopathic sensorineural hearing impairment by massively parallel sequencing. Y.H. Lin1,2, C.C. Wu1,3, T.Y. Hsu1, W.Y. Chiu1, C.I. Hsu1, P.L. Chen1,2,3,6, 1) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 2) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 3) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 4) Department of Otolaryngology, E-Da Hospital, I-Shou University, Kaohsiung, Taiwan; 5) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 6) Graduate Institute of Medical Genetics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan.

Hearing impairment is a genetically heterogeneous condition with > 100 deafness genes identified thus far. Recent studies have confirmed the utility of massively parallel sequencing (MPS) in addressing the genetically heterogeneous hereditary hearing impairment. In our previous study, we used Illumina HiSeq2000 to sequence 80 genes in 12 multiplex families, and identified 2 known mutations and 2 novel mutations as the causative mutations in 4 families. Recently, we further upgraded the screening panel to include 131 deafness genes, and applied the panel to another 12 unrelated multiplex families with idiopathic sensorineural hearing impairment. Criteria for data filtering included: allele frequencies <5%, both PolyPhen2 and SIFT scores >0.95, Sanger sequencing, segregation pattern, and evolutionary conservation of amino acid residues. In the proband of an autosomal dominant family with 9 affected members, we found a novel single nucleotide deletion mutation, c.149delT, in exon 2 of GATA3. It translated, c.149delT should result in a frameshift with a premature stop codon after a new amino acid sequence (p.Tyr504Tyr). C.G. (a 54 year old deceased patient), c.149delT was identified in all the 9 affected family members and 1 family member with normal hearing; the absence of hearing phenotypes in the latter might be attributed to his young age. GATA3 haploinsufficiency is thought to be associated with HDR syndrome which is characterized by hypoparathyroidism, deafness, and renal anomalies. We examined the clinical features in 7 affected members in the family. All the 7 subjects revealed mild to moderate sensorineural hearing impairment. By contrast, except for 1 subject with bilateral palatal retraction, other subjects with the above-mentioned signs did not present any signs of hypoparathyroidism or renal anomalies, indicating relatively mild phenotypes in this family. To our knowledge, the present study represents the first report to achieve genetic diagnosis before the clinical diagnosis of HDR syndrome. We re-classified the clinical diagnosis in 2 initial clinical diagnoses. We re-visited patients in the latter case and found novel mutations. Of these 72 patients, 58 carried mutations in known LCA genes and 14 carried mutations in disease genes with de novo mode of inheritance. After the detailed research, only one variant p.E368K, in DNM2 gene was already reported as the cause of centronuclear myopathy. p.E368K variant was found in 43 variants were revealed with de novo mode of inheritance. After the detailed research, only one variant p.E368K, in DNM2 gene was already reported as the cause of centronuclear myopathy. p.E368K variant was found in 43 variants were revealed with de novo mode of inheritance. After the detailed research, only one variant p.E368K, in DNM2 gene was already reported as the cause of centronuclear myopathy. p.E368K variant was found in 43 variants were revealed with de novo mode of inheritance. After the detailed research, only one variant p.E368K, in DNM2 gene was already reported as the cause of centronuclear myopathy. p.E368K variant was found in 43 variants were revealed with de novo mode of inheritance. After the detailed research, only one variant p.E368K, in DNM2 gene was already reported as the cause of centronuclear myopathy. p.E368K variant was found in 43 variants were revealed with de novo mode of inheritance. After the detailed research, only one variant p.E368K, in DNM2 gene was already reported as the cause of centronuclear myopathy. p.E368K variant was found in 43 variants were revealed with de novo mode of inheritance. After the detailed research, only one variant p.E368K, in DNM2 gene was already reported as the cause of centronuclear myopathy.
3096T

**De novo** frameshift mutation in ASXL3 in a patient with global developmental delay, microcephaly, and craniofacial anomalies. D.L. Dinwoodie1,2, S.E. Sedeno1,2, C.J. Saunders1,2, N.A. Miller3,4, E.G. Farrow1,2, L.D. Smith1,2, S.F. Kingsmore1,2,6.

1) Department of Pediatrics, University of New Mexico Health Sciences Center, Albuquerque, NM; 2) Clinical Translational Science Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico; 3) Center for Pediatric Genomic Medicine, Children’s Mercy Hospital, Kansas City, Missouri; 4) Department of Pediatrics, Children’s Mercy Hospital, Kansas City, Missouri; 5) Department of Pathology, Children’s Mercy Hospital, Kansas City, Missouri; 6) School of Medicine, University of Missouri-Kansas City, Kansas City, Missouri.

Currently, diagnosis of affected individuals with rare genetic disorders can be lengthy and costly, resulting in a diagnostic odyssey and in many patients a definitive molecular diagnosis is never achieved despite extensive clinical investigation. The recent advent and use of genomic medicine has resulted in a paradigm shift in the clinical molecular genetics of rare diseases and has provided insight into the causes of numerous rare genetic conditions. In particular, whole exome and genome sequencing of families has been particularly useful in discovering *de novo* germline mutations as the cause of both rare diseases and complex disorders. We present a six year old, African American female with microcephaly, autism, global developmental delay, and metopic craniosynostosis, who is nonverbal, and suffered intrauterine growth restriction. Exome sequencing of the patient and her two parents revealed a heterozygous two base pair *de novo* deletion, c.1897_1898delCA, p.Gln633ValfsX13 in the ASXL3 gene, predicted to result in a frameshift at codon 633 with substitution of a valine for a glutamine and suffered intrauterine growth restriction. By PolyPhen2. Phenotypic evaluation of the samples with these rare, non-synonymous variants revealed zero patients with developmental delay, microcephaly, or other craniofacial anomalies, suggesting that these variants are unlikely to be pathogenic in a heterozygous state. In conclusion, we expand the knowledge about disease causing mutations and the genotype-phenotype relationships in the ASXL3 gene and provide evidence that rare, non-synonymous, damaging mutations are not associated with developmental delay or microcephaly.

3097F

Unraveling disease genes causing autosomal recessive disorders in Qatari population by whole exome sequencing. S. Fahimi


1) Department of Human Genetics, Faculty of medicine, McGill University and Genome Quebec Innovation Center, Montreal, Quebec, Canada; 2) Department of Clinical and Metabolic Genetics, Department of Pediatrics, Hamad Medical Corporation, Doha, Qatar; 3) Section of Cytogenetic, Department of Pathology and Laboratory Medicine, Hamad Medical Corporation, Doha, Qatar.

Background: Whole Exome Sequencing (WES) was applied as a molecular diagnostic tool to identify disease-causing mutations in autosomal recessive disorders (ARD) in Qatari population where alternative molecular diagnostic tools had failed to detect pathogenic variants in causal genes. ARD are usually severe and rare, resulting from the transmission of one defective allele by each parent, which occurs with higher rate in consanguineous families. The prevalence of consanguineous marriage is high in Qatar (47%) that results in a higher incidence of several ARD. To reduce the overall socio-economic burden of such diseases, the development of diagnostic tools and prevention strategies is a priority for our population. To achieve these goals, the genes causing human genetic diseases should be first discovered. Methods: Whole exome capturing, sequencing and bioinformatics analyses were performed using our standard protocols at Genome Quebec Innovation Center, Canada. Results: WES was performed on 26 consanguineous Qatari families. The mode of inheritance was assumed to be AR because of unaffected status of parents, consanguinity and having equally affected male and female children. In families with one or only male affected child, *de novo* and X-linked inheritance were also considered. This led to the identification of definitive causal mutations in 12 families: Hypophosphatemic rickets; Hurler syndrome; Glycogen storage disease; Noonan-Like syndrome; Seckel syndrome; Geleophysic dysplasia; Limb-girdle muscular dystrophy; Multiple Fractures; Metachromatic Leukodystrophy; Immunodeficiency and Juvenile onset cataract. For 6 other families, we identified several candidate (1-26) genes in which the validation or functional studies are in progress: mental retardation; CNS anomaly; eye anomalies; peripheral neuropathy; axonal peripheral neuropathy and Oro-facio-digital syndrome. The 8 remaining families are still in progress. Furthermore, with decreasing sequencing costs and improving analysis pipelines, we expect WES to be in widespread clinical use in the near future that will help us to decrease time and cost of diagnosis and to focus on appropriate treatment and supportive care.
Exome sequencing unveils novel disease-causing variation in a Charcot-Marie-Tooth disease cohort. C. Gonzaga-Jauregui1, T. Hare1, D. Pehlivan1, Y. Okamoto1, W. Wisniewski1, D. Muzny2, R.A. Gibbs1, J.R. Lupski1, 2, 3. Centers for Mendelian Genomics. 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX. Charcot-Marie-Tooth (CMT) disease is the most common hereditary neuropathy affecting approximately 1/2500 individuals. CMT is a clinically heterogeneous distal symmetric polyneuropathy (DSP) with two major groups distinguished electrophysiologically: demyelinating CMT1 and axonal CMT2. Although the major cause of CMT, the 1.4 Mb duplication CNV of 17p11.2 that is responsible for ~70% of the CMT1 cases, is used extensively in clinical diagnosis and part of the evidenced based practice guidelines in the United States for evaluation of DSP, CMT shows extensive underlying genetic heterogeneity with about 50 loci identified or linked to date to the different subtypes of the disease. Furthermore, homozygous CMT1 duplication CNV and triplication CNV can portend a more severe clinical course while each conveys distinct genetic recurrence risk. Exome sequencing allows assessing all the coding variation in the 2% of the human diploid genome that we can interpret through our knowledge of the genetic code; however, personal exomes contain ~10,000 - 12,000 nonsynonymous variants and many of these are novel. Further, even in genetic conditions with known responsible genes, interpretation can be complicated by the presence of novel variants in more than one causative gene. We have performed exome sequencing at high coverage (> 100x) of a cohort of 39 patients with different clinical presentations of CMT, in whom the genetic cause had not previously been identified using a multitude of clinically available and research molecular genetic analyses. We have found the apparent causative mutations in several individuals, and potentially disease causing mutations in novel genes in additional subjects. We also show that affected individuals can have known or novel rare variants in multiple CMT genes, possibly contributing to the mutational load and phenotypic variation in the disease. These findings are consistent with the recently proposed Clan Genomics hypothesis [Cell (2011);147(1):32-43] which posits that new mutations in novel genes in additional subjects. We also show that affected individuals in the family or clan, and novel combinations from the proband's ancestors may be responsible for the disease. These findings are consistent with the recently proposed Clan Genomics hypothesis [Cell (2011);147(1):32-43] which posits that new mutations in novel genes in additional subjects. We also show that affected individuals in the family or clan, and novel combinations from the proband's ancestors may be responsible for the disease.
542

Posters: Molecular Basis of Mendelian Disorders

3101W

Functional Evaluation of candidate mutations identified in whole exome sequences of patients with undiagnosed diseases. Y. Lu1, P. Xie 1, E. Ruzzo 1, A. Need2, V. Shashidhar 2, Y. Jiang3, X. Zhu4, D. Goldstein5, 1) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina, USA; 2) Department of Pediatrics, Duke University School of Medicine, Durham, North Carolina, USA.

A number of recent studies indicate that whole exome sequencing of both patients and parents (trio sequencing) can resolve up to 50% of undiagnosed childhood diseases (Need et al. 2012). We have recently sequenced 15 such trios and evaluated functional consequences of candidate mutations. Here we report functional consequences of three mutations of particular interest. A proband with severe osteopetrosis, autism, and developmental delay, carries a homozygous nonsynonymous mutation, R95H, in sorting nexin 10 (SNX10). SNX10 is essential for formation and repositioning of the mitochondria which function will result in osteopetrosis. Our data from in vitro analysis suggests that this mutation leads to loss of function of SNX10. Another hemizygous nonsynonymous mutation (rs148886271; P670S) in BCL6 corepressor (BCOR) was found to down-regulate protein expression in a well-characterized 1721 amino acid (a.a) isoform, but induced a greater than 50-fold up-regulation in another abundant a.a. isoform whose function is uncharacterized. The evidence together suggests a complicated influence on normal BCOR function as a transcriptional factor co-regulator and possibly leads to altered downstream transcriptional activation that may cause global developmental delay observed in this patient. Several other genes which are of great interests but still lack functional evidence include ANK3, which has been implicated as a candidate gene in epilepsy and autism. Compound heterozygous mutations in ANK3 were found in one patient with severe intractable seizures in our cohort. We are planning to proceed functional evaluation of ANK3 through mouse hippocampus primary culture. Further studies, including confirming molecular results by patients RNA and protein samples are still required to fully understand the influence of mutations in ANK3. Our findings emphasize the central role of functional evaluation in the interpretation of personal genomes, since many of the most interesting mutations have not been characterized before. Our work also indicates that patients with undiagnosed diseases may carry mutations in genes which are of great interests but still lack functional evidence.

3102T

Whole exome sequencing identifies novel mutations in three families with GAPO syndrome. D. Palhivan1, E. Karaca 1, T. Gambin1, S.N. Jhangiani2, G. Gonzaga-Jauregui 1, M. Selman Yildirim 1,2, D. Muzny1, M.M. Atik 1, R.A. Gibbs 1, S. Selman Yildirim 2, A. Zamani 1, D. Guly1, N.H. Elcioglu 1, B. Bozkurt 1, J.R. Lupski3, 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Neurogenetics and Functional Genomics, Cleveland Clinic, Cleveland, OH.

GAPO syndrome (MIM# 230740) is the acronym for Growth retardation, Alopeica, Psuedoanodontia, and Optic atrophy. So far about 30 cases have been reported worldwide, making it one of the rarest recessive conditions. Distinctive craniofacial features including alopecia, rarefaction of eyebrow Bones and eyelashes, frontal bossing, high forehead, midfacial hypoplasia, hyperelorism, and thickened eyelids and lips make it a readily recognizable phenotype. However there are additional clinical findings in each reported case. Mutations in ANTXR1 has recently been found to be causative of the GAPO syndrome. ANTXR1 plays a role in the regulation of extracellular matrix. In this study we applied whole exome sequencing in three GAPO syndrome families with six affected individuals. Exome sequencing analysis revealed two novel mutations: one frameshift, one splice site, and previously described mutations in the ANTXR1 gene. We present clinical findings of patients with GAPO syndrome. Our patients contribute to make a better genotype-phenotype correlation in this rare condition and are important to understand the role of ANTXR1 in the function/dysfunction of the extracellular matrix.

3103F

A genome-wide catalogue of genetic variants for nephrotic syndrome via whole genome sequencing. M.G. Sampson1,2, A. Tam2, C.A. Gadegbeku- Sanders3,4, M. Kretzler1,5, H.M. Kang,2, Nephrotic Syndrome Study Network (NEPTUNE), The Michigan O’Brien Renal Center (CPROBE), 1) Pediatrics and Communicable Disease; 2) Public Health-Biostatistics; 3) Internal Medicine-Nephrology; 4) University of Michigan, Ann Arbor, MI; 5) Temple University, Philadelphia, PA; 6) Case Western Reserve University, Cleveland, OH.

Nephrotic syndrome (NS) is a rare glomerular condition with more than 20 known Mendelian genes and few common risk variants. Yet, these explain only a subset of the population prevalence observed across the clinical spectrum of NS, particularly in adults. To identify known and novel variants associated with NS molecular, histologic, and clinical phenotypes, we are performing low-pass whole genome sequencing (WGS) and exome chip genotyping on 550 affected subjects recruited into observational studies on NS from four US nephrology centers. Here we present initial data on the first 256 patients. Subjects were either recruited at time of initial indicated biopsy for suspected primary NS or had an existing diagnosis of focal segmental glomerulosclerosis (FSGS) or membranous nephropathy (MN). All underwent Illumina Exome Chip genotyping and Illumina HiSeq WGS. WGS data underwent GlioCloud pipeline with linkage disequilibrium refinement, and comparison to Exome Chip genotypes. The variant frequencies were compared to the latest release of 1000 Genomes and Exome Sequencing Projects, stratified by ancestry. Diagnoses were 52% minimal change disease/FSGS, 25% MN, and 24% other glomerulopathy. Subject ancestry was 49% European (EUR), 29% African (AFR), 15% Admixed American (AMR), and 7% Asian (ASN). Mean sequencing depth was 4.5x. 20.9M SNPs were identified, with transition to transversion (Ts/Tv) ratio 1.7, and 16% of SNPs were novel to dbSNP. The X chromosome was excluded. Using the most frequent 1.6M SNPs, we predicted gene sets using Ingenuity Pathway Analysis. We identified 17 genes that were significantly enriched by SNPs associated with NS, including CREBBP, NFKB1, PIK3CA, and SMAD4. In 49% of patients, we found one or more rare NS-associated variants. Overall, 10% of patients carried one or more rare NS-associated variants. Finally, we found no evidence of a rare causal variant but our sample size may have been insufficient to detect small effect variants.

3104W

Personalized functional genomics elucidates novel Mendelian disease genes and provides proof of pathogenicity for variants of uncertain significance. P. Bonnen1,2, A. Besse1,2, S. Lalani3, W. Craigien4, F. Scaglia5, R. McFarland6, C. Bacio1, R. Taylor2, K. Scott2, 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Welcome Trust Centre for Mitochondrial Research, The Medical School, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

Genome-wide, the foremost efforts of our efforts to identify genetic underpinnings of disease. Bioinformatic analyses of sequence data alone can only predict the impact of variants in known disease genes. For discovery of novel disease genes and to provide proof of pathogenicity of specific alleles, we combine genome-wide sequencing efforts with robust pipelines that allow functional characterization of gene variants in patient cells. This scalable infrastructure is intended to accelerate validation of mutations driving single gene disorders with the ultimate goal of translating these findings into clinical diagnostics and therapeutics. Combining the resolution of genome-wide sequencing with functional profiling of variants in patient cells enables the personalized identification of bona fide pathogenic mutations. Our approach combines whole exome sequencing with (1) consistent, systematic and thorough functional profiling on every patient regardless of variation in clinical presentation, (2) bioinformatic analyses to prioritize pathogenic variants, (3) delivery of the wild-type copy of suspected pathogenic genes to rescue cellular dysfunction and (4) delivery of patient mutations with simultaneous knockdown of the endogenous gene of interest into healthy cells to recapitulate cellular disease phenotypes. Our approach has already revealed several novel disease genes, and importantly, has enabled validation of the pathogenicity of alleles reported as variants of uncertain significance (VUS) in the diagnostic setting. Our current study cohort consists of children with multi-system disease with clinical and diagnostic assessment that present with mitochondrial, neuromuscular, neurologic, cardiomyopathy, bone, and rapid aging diseases. Our personalized functional genomics approach has discovered novel disease genes and validated pathogenic mutations in greater than 60% of alleles reported as VUS. This approach serves as a model for validation of sequence-based genetic findings, and our results provide new insights into the mechanisms underlying the pathology of these single gene disorders and delivers genes and pathways for further study of the pathogenic mechanisms underlying diseases with more complex etiology.
3105T Genotype-phenotype correlation in a national mutation study of Danish patients with HHT. P.M. Torring1,2, K. Brugaard1, L.B. Ousager1, P.E. Andersen3, A.D. Kjeldsen4, HHT centre, Odense University Hospital, 1; Dept. of Clinical Genetics, Odense University Hospital, Odense, Denmark; 2 Dept. of Otorhinolaryngology, Odense University Hospital, Odense, Denmark; 3 Interventional Radiology, Odense University Hospital, Odense, Denmark.

Purpose: Hereditary Haemorrhagic Telangiectasia (HHT) is an autosomal dominantly inherited vascular disease characterized by the presence of mucocutaneous telangiectasia and visceral arteriovenous malformations (AVM). About 85% of HHT patients carry mutations in the ENG, ACVRL1 or SMAD4 genes. Here, we report on the genetic heterogeneity in the Danish national HHT population and address the prevalence of pulmonary AVM (PAVM). Methods: Proband of 107 apparently unrelated families received genetic testing, including sequencing and multiplex ligation-dependent probe amplification (MLPA) analyses of ENG, ACVRL1 and SMAD4. Results: In 89% of the probands (n=95), a mutation was identified in one of the three genes. We identified 64 unique mutations, primarily in ENG and ACVRL1, of which 27 (41%) were novel. Large deletions were identified in both ENG and ACVRL1. The prevalence of PAVM was 52.3% in patients with an ENG mutation and 12.9% in the ACVRL1 mutation carriers. We diagnosed 80% of the patients clinically, fulfilling the Curaçao criteria, and those remaining were diagnosed by genetic testing. Conclusions: Proper genetic testing requires analysis for both mutations and large rearrangements of all three genes. Inclusion of a pathogenic mutation as a diagnostic criterion is discussed.

3106F Exome sequencing of subjects with Congenital Insensitivity to Pain and their family members to reveal novel pain genes. B. Zhang1, W. He1, J. Stephens1, B. Sidders2, S. Scollen2, S. Paciga3, L. Wood1, M. Sudworth4, G. Johnson3, S. John3, L. Danziger3, S. Vangeli4, 1 Pfizer Inc, Eastern Point Road, Groton, CT 06340, USA; 2 Pfizer Neustadt, Granta Park, Great Abington, CB21 6GS, UK; 3 Pfizer Ltd, Ramsgate Rd, Sandwich, CT13 9NJ, UK; 4 Department of Clinical Neurophysiology and Pain Center, Groupe Hospitalier Pitié-Salpêtrière, Paris, France.

Congenital insensitivity to pain (CIP) is a rare heterogeneous condition comprised of an absence of sensation to noxious stimuli and can be accompanied by a range of other clinical abnormalities, including anosmia, anhidrosis, and mental retardation. Three genes have consistently been shown to harbour mutations in patients with CIP - SCN9A, NTRK1, and NGF - but some studies that have sequenced these genes failed to identify mutations segregating in families with the disorder. It is therefore likely that there are mutations in other genes responsible for CIP, which may represent novel therapeutic targets for pain. We sequenced the exome of 30 subjects - 15 with CIP and 15 unaffected family members - using the Agilent SureSelect v4 all exon 51Mb kit for exome capture and the Illumina HiSeq 2000 for sequencing. Sequences were aligned using BWA and variants called using GATK. Depending on the mode of inheritance and the levels of relatedness within the families, different mutation filtering strategies were employed. Mutations were filtered using frequency cutoffs in public databases including 1000 genomes, dbSNP and the exome variant server, as well as from 400 in-house exome sequenced samples. Autozygosity mapping was used to identify genomic regions of particular interest. Of the 10 families, 2 of them have potentially functional mutations that segregate in known genes - one family has a novel nonsense mutation in SCN9A and another has a 27bp deletion in the NTRK1 gene. For the other 8 families, we have identified lists of segregating mutations, but of these, the causative mutations remains to be elucidated. Future steps include sequencing coding regions that had poor coverage as well as putative regulatory regions of the 3 known CIP genes, as well as GWAS chip genotyping of all family samples to more accurately define runs of homozygosity and/or regions identical by descent. It is likely that the CIP in several of the families will not be the result of known CIP genes, and thus these future steps are likely to reveal novel genes responsible for CIP.

3107W Whole exome sequencing of a dominant retinitis pigmentosa family with female-specific expressivity identifies a novel deletion in PRPF31. E. Chang1,2, W. Chang1,2, S. Aso1,2, M. Villasenov2, K. Hart1, R.W. Miller1, E.T. Demirtasakis1, N. Katsanis1, Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) Virtual Eye Care MD, Merida, Yucatan, Mexico; 3) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland.

Photoreceptor dysfunction is the most common cause of visual impairment and affects ~1 in 3,000 humans. Underscored by an excess of 180 primary disease gene loci, inherited retinal disorders are hallmark by vast genetic heterogeneity. Understanding the molecular basis of these diseases requires analysis for both mutations and large rearrangements of all three genes. Inclusion of a pathogenic mutation as a diagnostic criterion is discussed. Conclusions: Proper genetic testing requires analysis for both mutations and large rearrangements of all three genes. Inclusion of a pathogenic mutation as a diagnostic criterion is discussed.

3108T A new palmoplantar keratoderma with severe erythermalgia alleric to Olmsted syndrome. S. Duchatelet1, S. Faveur2,3, S. De Veer4, S. Fraipont2, P. Nitschke2, C. Bole-Feyssy3, C. Bodeney2, A. Hovnanian1,2,6,7


We describe a new severely affected family: a seven-year-old girl from healthy and non consanguineous French parents. The patient presented with progressive and severe palmoplantar keratoderma (PPK) associated with intense erythermalgia manifesting by acute flares of inflammation, redness, burning and pain. The lesions were located in the soles and palms, with or without hyperkeratosis. The dysmorphology of the hands (hands, feet, and nails) was demonstrated using a dermatoscope. Microscopic examination shows superficial irregularities without specific abnormalities under polarized microscopy. Whole exome sequencing identified a de novo heterozygous p.Leu673Pro mutation within TRPV3 (transient receptor potential cation channel, subfamily V, member 3) encoding a nonselective cation channel involved in a variety of processes, including temperature sensation and vasoregulation. Recently, distinct missense TRPV3 mutations were identified in Olmsted syndrome (OS), a rare keratinizing disorder with peripheral hyperkeratosis and multilayer PPK, all features which were absent in our patient. In addition, the association of severe keratoderma clearly distinguishes our patient from previously described cases of PPK. The p.Leu673Pro missense mutation involves a highly conserved amino acid residue across species and is predicted to be damaging by in silico analysis. Thus, this mutation introduces a significant perturbation in a critical position of the molecule. Using a TRPV3 homology model, we show that p.Leu673 is located immediately above the predicted activation gate residue Met677 on the preceding S6 transmembrane helical turn. We suggest that this mutation impairs channel assembly and causes severe palmoplantar keratoderma phenotype associated with TRPV3 mutations. The molecular mechanisms through which TRPV3 mutations lead to OS or other palmoplantar keratoderma phenotype are yet to be explored.
3109F
A mutation in A-band titin is associated with hereditary myopathy with early respiratory failure in a Japanese family. R. Izumi, Y. Niinomori, Y. Aoki, N. Suzuki, M. Kato, H. Wartez, T. Takahashi, M. Tateyama, T. Nagashima, K. Funahama, K. Nakayama, M. Aoki, Y. Matushara, 1) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan; 2) Department of Neurology, Tohoku University School of Medicine, Sendai, Japan; 3) Department of Neurology and Division of Clinical Research, National Hospital Organization Nishitaga National Hospital, Sendai, Japan; 4) Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 5) Department of Neurology, Okayama University Medical School, Okayama, Japan.

Hereditary myopathy with early respiratory failure (HMNERF, MIM #603689) was originally described as an autosomal dominant disease characterized by adult onset proximal or distal myopathy with early respiratory failure and overlapping pathologic findings with myofibrillar myopathy. HMNERF had been considered an extremely rare disease, caused by a rare TTN gene mutation (c.97348C>T of the kinase domain) among North European populations. In 1993, our group reported a Japanese family with dominantly inherited cytoplasmic body myopathy, which is now included in myofibrillar myopathy. Currently, this family includes 20 patients in five successive generations who show almost homogeneous clinical features characterized by chronic progressive distal muscle weakness and early respiratory failure. However, the underlying genetic etiology in this family was unknown. In this study, we performed linkage analysis and whole exome sequencing to identify the responsible genetic mutation of the family, clinically and pathologically compatible with HMNERF and identified a novel c.90263G>T mutation in the A150 domain. This mutation is compatible with HMNERF and identified a novel c.90263G>T mutation in the A150 domain, which has significant roles in the pathogenesis of HMNERF as well as the kinase domain, although detailed mechanisms of the pathogenesis remain unknown. HMNERF is a more frequent disease than previously described in a Finnish family.

We have performed exome capture and sequencing of 33 individuals with HD (including 10 affected sib-pair) and where available an unaffected sibling and the unaffected parents (n=18) from multiple kindreds with only Hodgkin's lymphoma. The family structure of majority of these HD sib-pairs reflects a de novo or incomplete mode of inheritance. Targeted exome sequencing was performed using Agilent SureSelect 30MB and sequencing was achieved using Illumina HiSeq chemistry. Mapping and variant calling were performed using BWA and GATK either at Broad Institute or at MSKCC. QC was performed using variant annotation using SNPEff. Variants were filtered based on public control data such as ESP6500, 1000genomes, prediction algorithms such as SIFT and PolyPhen2 and on conservation scores such as GERP and PhyloP. Analyses were performed in both recessive and dominant mode where applicable. Recurrence of variant analyses was also performed.

We did not observe truncating NPAT mutations in any family. We also did not observe any segregating high quality, high impact mutations in REL, PVT1 or GATA3. No such mutations were observed in the HLA gene cluster.

Ongoing efforts are geared to the discovery of segregating rare variants in the germline including genes reportedly mutated in somatic tissue in this type of lymphoma. To understand the population risk and the prevalence, the frequency of these genic-variants need to be estimated in a larger cohort of familial lymphomas. Future plans include a targeted capture of prioritized genes in 1000 sporadic HD cases ascertained at MSKCC.

Acknowledgement: Geoffrey Beene cancer research fund, the Starr Cancer Consortium and the Lymphoma Foundation.

3110W
Exome Sequencing of Familial Hodgkin’s Kindreds. V. Joseph, T. Thomas, A. Artemov, K. Schrader, A. Kleizen, C. Marschreck, R. Maury, M. Correia, X. Wei, N. Gupta, M. Margolin, A. Zelenetz, C. Portlock, G. Getz, R. Klein, D. Haber, M. Daly, S. Lipkin, D. Altshuler, K. Offit, 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Cancer Genome Computational Analysis, Broad Institute, Boston, MA; 5) Weill Cornell Medical College, New York, NY; 6) Genomics Platform, Broad Institute, Boston, MA; 7) Lymphoma Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 8) Massachusetts General Hospital Cancer Center, Boston, MA; 9) Medical & Population Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY; 10) Department of Neurology, Memorial Sloan-Kettering Cancer Center, New York, NY; 11) Department of Genetics, Medical College of Wisconsin, Milwaukee, WI.

Hodgkin’s disease (HD) is a cancer of the immune system characterized by the presence of Reed-Sternberg cells. Concordance for HD seen in identical twins and increased familial aggregation suggests that there are genetic susceptibility factors involved in its predisposition. Genome-wide association studies have shown there are multiple loci associated with HD at chromosome 6p21 HLA locus and several other loci near genes such as REL, PVT1 and GATA3. A germline mutation in NPAT in a family was described in a Finnish family.

3111T
ITPR2 loss-of-function mutation causes familial generalized anhidrosis and hyperthermia. J. Klar, C. Hisatsune, S. M. Baig, M. Tanqi, A. C. V. Johansson, M. L. Foss, N. Naved, K. Sugiuara, 1) Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden; 2) The Laboratory for Developmental Neurobiology, RIKEN Brain Science Institute, Saitama 351-0198, Japan; 3) The Human Molecular Genetics Laboratory, National Institute for Biotechnology and Genetic Engineering (NIBGE), 38000 Faisalabad, Pakistan; 4) The Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

Anhidrosis, defined as the absence of perspiration in the presence of an appropriate stimulus such as heat or exercise, is a rare condition that may be acquired or congenital. The reasons for anhidrosis, or reduced sweating (hypohidrosis), are heterogeneous and may be caused by defects of sweat gland innervation in disorders of the autonomous nervous system or by a reduction of functional sweat glands in different ectodermal syndromes. Reports on Familial generalized anhidrosis with normal sweat glands (GANSG) (MIM 106190) are very few.

We have identified a consanguineous Pakistani kindred segregating autosomal recessive GANSG in five affected family members. Using gene mapping and targeted re-sequencing we identified a novel homozygous missense mutation (c.7492G>A; p.G2498S) in the ITPR2 gene associated with anhidrosis. The ITPR2 gene encodes for the Insitol 1,4,5-trisphosphate receptor type 2 (InsP3R2), a member of the Inps3r protein family of intracellular Ca2+ channels. Functional studies showed that the mutation affects the selective filter for Ca2+-ions with a resulting loss of function of InsP3R2. Furthermore, the InsP3R2-/- mice show a markedly reduced sweat gland response upon pilocarpine stimulation when compared to wild-type littermates. Taken together, our findings indicate that the ITPR2 mutation p.G2498S underlies anhidrosis in the patients investigated. Furthermore, we have identified InsP3R2 mediated Ca2+ release as a critical mechanism for eccrine sweat gland function, perspiration and thermal cooling.
3112F A comprehensive disease-mutation search of mitochondrial respiratory chain disorder. M. KOHDA1, Y. Tokuzawa2, Y. Moriyama3, H. Kato4, Y. Kishita2, N. Uehara2, S. Tamaru4, Y. Yamashita-Sugahara5, Y. Nakachi5, N. Matoba1, T. Yamazaki5, M. Mori4, K. Murayama5, Y. Mizuno4, A. Ohtake6, Y. Okazaki7,2,1 1) Div Translational Res, Research Center for Genomic Medicine, Saitama Med Univ, Hidaka, Saitama, Japan; 2) Div of Functional Genomics & Systems Medicine, Research Center for Genomics, Saitama Medical University, Hidaka, Saitama, Japan; 3) Div. of Developmental Biology, Research Center for Genomic Medicine, Saitama Medical University, Hidaka, Saitama, Japan; 4) Dept of Obstetrics and Gynecology, Saitama Medical University Hospital, Moroyama, Saitama, Japan; 5) Dept. of Pediatrics, Saitama Medical University, Moroyama, Saitama, Japan; 6) Dept. of Pediatrics, Jichi Medical University, Tochigi, Japan; 7) Dept. of Metabolism, Chiba Children’s Hospital, Chiba, Japan.

Mitochondria are small organelles and serve as the powerhouse of the living cells because they generate vital energy in their respiratory chain system consisting of protein complexes I to V. Mitochondrial respiratory chain disorder (MRCD) is an intractable disease that develops in childhood. It is a highly frequent inborn error of metabolism that occurs in at least one out of every 7,000 births. Prominent symptoms develop in such organs as the brain, heart, liver and muscles, where a great deal of energy is required. In most cases, effective treatment has not been established yet. It is known that various gene abnormalities cause defects of the protein complexes of the respiratory chain, which results in mitochondrial dysfunction. However, identification of the causative gene and the understanding of pathogenic mechanism of MRCD remain largely unsolved. In this study, we applied SNP array and exome sequencing in combination with stepwise filtering of gene variants. Exome sequencing data were filtered using three different criteria: (i) the presence of mutations in known disease causing genes; (ii) mutations in genes that code for mitochondrial proteins; (iii) unbiased genome-wide approach with strict filtering strategy. For this study, 103 unrelated individuals were chosen who display juvenile-onset mitochondrial disorders. In 18/103 cases, we identified mutations in known disease-causing genes (AARS2, ACAD9, BOLA3, COX10, GFM1, MPV17, NDUFA1, NDUFA10, NDUFAF6, PC, RARS2, SUCLA2, SURF1 and TUFM). Currently we prioritize the most promising candidate variants for further filtering and to identify rare variants. We are evaluating the expression of the candidate genes in fetal and adult retina in our global expression databases and will be validating significant causal variants using Sanger sequencing. Initial analysis of three families has led to the identification of rare variants in CDHR1, C8orf18 and GUCY2D genes. CDHR1 and C8orf18 were previously reported to cause AR form of CRD and GUCY2D was previously implicated in the cause of AD CRD. In CDHR1 we have identified a new non-sense mutation predicted to result in a truncated protein with loss of several important domains. Further experiments will be carried out to understand the biological effect of the novel variant and the causative disease pathology. Our studies should help us in better understanding the molecular basis of cone and cone-rod dystrophy, thereby improving clinical management and development of new therapeutic strategies.

3113W Mutation in ATP6AP2, an essential accessory subunit of vacuolar ATPase, causes X-linked Parkinson Disease with Spasticity (XPDS), E. Korvatka1, T. Strovas2, D.-H. Chen2,3,5, B.C. Kraemer2,3,5, Y. Moriyama3, H. Kato4, Y. Kishita2, N. Uehara2, S. Tamaru4, Y. Yamashita-Sugahara5, Y. Nakachi5, N. Matoba1, T. Yamazaki5, M. Mori4, K. Murayama5, Y. Mizuno4, A. Ohtake6, Y. Okazaki7,2,1 1) Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 2) Geriatric Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle; 3) Department of Neurology, University of Washington, Seattle; 4) VISP-20 Mental Illness Research, Education, and Clinical Center, Department of Veteran Affairs, Seattle; 5) Parkinson’s Disease Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle; 6) Department of Immunology, University of Washington, Seattle; 7) Department of Genome Sciences, University of Washington, Seattle; 8) Department of Medicine (Medical Genetics), University of Washington, Seattle; 9) Department of Medicine (Gerontology Division), University of Washington, Seattle.

We report a novel gene for a parkinsonian disorder. X-linked Parkinson Disease with Spasticity (XPDS) presents either as typical adult onset Parkinson’s disease or earlier onset spasticity followed by parkinsonism. We previously mapped the XPDS gene to a 28 Mb region on Xp11.2-X13.3. Exome sequencing of one affected individual identified five rare variants in this region, of which none was missense, nonsense or frame shift. Using patient-derived cells we tested the effect of these variants on expression/splicing of the relevant genes. A synonymous variant in ATP6BP2, c.345C>T (p.S115S), markedly increased exon 4 skipping resulting in overexpression of a minor splice isoform that produces a protein with internal deletion of 32 aa in up to 50% of the total pool, with concomitant reduction of isoforms containing exon 4. ATP6BP2 is an essential accessory component of the vacuolar ATPase responsible for lysosomal degradative functions and autophagy, a pathway frequently affected in Parkinson’s disease. Reduction of the full size ATP6AP2 transcript in XPDS cells and decreased level of ATP6AP2 protein in XPDS brain may compromise V-ATPase function, as seen with siRNA knockdown in HEK293 cells, and may ultimately be responsible for the pathology.

3114T Identification of genetic defects in cone and cone-rod dystrophy by whole exome sequencing. C. Lazar1,2, L. Zelling2, M. Mutsuddi1,2, D. A. Nickerson2,3,4,5, K. Kiianitsa1, Y. Tokuzawa2, Y. Moriyama3, H. Kato4, Y. Kishita2, N. Uehara2, S. Tamaru4, Y. Yamashita-Sugahara5, Y. Nakachi5, N. Matoba1, T. Yamazaki5, M. Mori4, K. Murayama5, Y. Mizuno4, A. Ohtake6, Y. Okazaki7,2,1 1) Neurobiology-Neurodegeneration & Repair Laboratory (N-NRL), National Eye Institute, National Institutes of Health, Bethesda, MD, 20892; 2) Institute for Doctoral Studies, Babes-Bolyai University, Mihal Kogalniceanu 1, Cluj-Napoca, Romania, 400084; 3) Department of Ophthalmology, Hadassah - Hebrew University Medical Center, Jerusalem, Israel 91120; 4) Department of Molecular and Human Genetics, Varanasi, India 221005.

Cone dystrophy (CD) and cone-rod dystrophy (CRD) are clinically and genetically heterogeneous retinal disorders displaying autosomal dominant (AD), autosomal recessive (AR) and X-linked inheritance patterns. Several genetic loci have been implicated, and causative mutations have been identified in over twenty genes. The goal of this study is to identify the cause of disease in AD and AR families with cone-dominated retinal phenotypes using whole exome sequencing. Finding causal variants in retinal dystrophies is difficult using traditional methods because of the high degree of genetic heterogeneity. Whole exome sequencing (WES) has been successfully applied for molecular characterization of cohorts affected with a heterogeneous condition as well as identification of novel variants. In this study, WES was performed on 13 individuals from four unrelated Israeli families. Three of the families were clinically diagnosed with CRD and one with CD. To capture the target regions from genomic libraries, we used the Agilent Sure Select Human All Exon V4 capture kit. Sequencing was carried out on an Illumina Genome Analyzer IIx. The sequencing reads that passed the initial quality control achieved an average coverage of 60X. Initial data analysis using Sanger sequencing identified three patients carrying likely deleterious variants in CHN2, NPHP3 and VPS13B. Subsequent filtering of the putative candidates in fetal and adult retina in our global expression databases and will be validating significant causal variants using Sanger sequencing. Initial analysis of three families has led to the identification of rare variants in CDHR1, C8orf18 and GUCY2D genes. CDHR1 and C8orf18 were previously reported to cause AR form of CRD and GUCY2D was previously implicated in the cause of AD CRD. In CDHR1 we have identified a new non-sense mutation predicted to result in a truncated protein with loss of several important domains. Further experiments will be carried out to understand the biological effect of the novel variant and the causative disease pathology. Our studies should help us in better understanding the molecular basis of cone and cone-rod dystrophy, thereby improving clinical management and development of new therapeutic strategies.
A novel germline PIGA mutation in Ferro-Cerebro-Cutaneous Syndrome: A neurodegenerative X-linked encephalopathy with epilepsy and systemic iron overload. R.L. Margraf1, E.M. Coonrod1, J. Uhlmann1, K. Mailempati1, A. Kumanovic1,2, T.M. Newcomb3, J.M. Opitz4,5,6, J.C. Carey4,6, H. Zhou2, B.E. Katz7, K.V. Voelkerding1,2, K.J. Swoboda1,2
1) ARUP Institute for Clinical and Experimental Pathology®, ARUP Laboratories, Salt Lake City, UT; 2) Department of Pathology, University of Utah, School of Medicine, Salt Lake City, Utah; 3) Department of Neurology, University of Utah, School of Medicine, Salt Lake City, Utah; 4) Department of Human Genetics, University of Utah, School of Medicine, Salt Lake City, Utah; 5) Department of Obstetrics and Gynecology, University of Utah, School of Medicine, Salt Lake City, Utah; 6) Department of Pediatrics (Division of Medical Genetics) University of Utah, School of Medicine, Salt Lake City, Utah; 7) Private Practice of Pediatrics and Pediatric Neurology (retired), Twin Falls, Idaho, 83303, USA.

This report presents a family with three males affected by Ferro-Cerebro-Cutaneous Syndrome, a novel X-linked syndrome of neurodegeneration, cutaneous abnormalities, and systemic iron overload. Linkage studies demonstrated a shared haplotype at Xp21.3-Xp22.2, a region spanning 18 megabases and containing more than 100 candidate genes. When candidate gene sequencing proved unsuccessful, exome sequencing was used to determine the causal variant. A heuristic filtering approach identified several rare mutations in the linkage region, but only the PIGA mutation segregated with disease in the family. The PIGA gene contains a germline three base pair in frame deletion (deleted positions chrX:15342943-15342945, p.110delILEu). The leucine residue that is deleted in the affected males is highly conserved from humans to zebrafish. The unaffected great-grandfather had the same X allele as all the affected males, but without the PIGA mutation. This indicates that the PIGA mutation arose de novo on the X allele from the great-grandfather to the grandmother who then passed the PIGA mutation on to several children, including an affected male. PIGA encodes an enzyme in the GPI anchor biosynthesis pathway. Clonal expansion of cells with somatic PIGA mutations causes paroxysmal nocturnal hemoglobinuria (PNH). In PNH, blood cells including erythrocytes are deficient in GPI anchor proteins, reducing expression of critical cell surface proteins and predisposing cells to complement-mediated lysis. Recently, a family with a germline PIGA mutation was reported where affected males had multiple congenital anomalies and severe neurologic impairment resulting in infantile lethality. In contrast, affected boys in this report were born without anomalies and were apparently neurologically normal prior to onset of seizures after 6 months of age, with 2 surviving to the second decade. An affected individual was tested for the presence of GPI anchor proteins in granulocytes, monocytes, and erythrocytes; and only granulocytes were found to be deficient in GPI anchored proteins. This could explain the milder phenotype seen in our patients compared to PIGA mutations that lead to erythrocyte lysis or early lethality. The novel PIGA mutation in this family likely caused a reduction in GPI-anchor protein cell surface expression in various cell types, resulting in the observed novel and complex phenotype involving the central nervous system, skin and iron metabolism.

Deciphering the genetic background of PEHO-like syndrome. M. Muona1,2,3, A. Laan2,4,5, A.-K. Anttonen1,4,5,6, M. Somer2, A. Palotie1,4,5, A.-E. Lehesjoki1,4,5, 1) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland; 3) Folkhalsan Institute of Genetics, University of Helsinki, Helsinki, Finland; 4) Haartman Institute, Department of Medical Genetics and Research Program’s Unit, Molecular Medicine, University of Helsinki, Helsinki, Finland; 5) Neuroscience Center, University of Helsinki, Helsinki, Finland; 6) Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 7) Norio Centre, Rinnekoti Foundation, Helsinki, Finland; 8) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK; 9) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

PEHO syndrome (progressive encephalopathy with edema, hypsarhythmia, and optic atrophy) is an infantile onset neurodegenerative disorder. The clinical presentation in Finnish patients, homozygous for a founder mutation (unpublished), is relatively uniform and includes hypotonia, infantile spasms with hypsarhythmic, profound psychomotor retardation, optic atrophy, and progressive brain atrophy originating in cerebellum. A significant number of patients manifest many of these features in the absence of the typical neuroradiological findings, or with no sign of progression. These patients remain without proper diagnosis but are often classified as PEHO-like. We aim to decipher the genetic basis of PEHO-like syndrome by using exome sequencing. We have selected 29 Finnish, mostly sporadic, PEHO-like patients, who have been excluded for the PEHO founder mutation. Exome sequencing was performed using Illumina HiSeq 2000 platform. We carried out variant filtering using strategies assuming two different patterns of inheritance: recessive and de novo. In the former we selected potentially deleterious variants - either homozygous or compound heterozygous - of essential splice site, nonsense, frameshift, and missense types with a 1000 genomes minor allele frequency below 1%. In the ‘de novo’ analysis we included only heterozygous, potentially damaging variants absent from the control databases. Analysis of the first 21 exomes revealed likely pathogenic hemizygous mutations in CDKL5, a known infantile epileptic encephalopathy gene, in two male patients. Capillary sequencing of the patients’ parents showed that the mutations occurred de novo. In an affected sibling we identified likely pathogenic compound heterozygous mutations in ABAT, previously linked to epileptic encephalopathies in a few cases. One patient had a three-amino-acid duplication previously reported as pathogenic in SPTAN1. Finally, we have also identified potentially pathogenic mutations in genes without a previous connection to encephalopathies. We are currently analysing the exomes of the remaining eight patients. To facilitate identification of heterogeneous de novo variants we have also sequenced the exomes of parents of seven patients. Our findings imply that PEHO-like syndrome is genetically highly heterogeneous. A subset of patients had mutations in previously established disease genes, indicating the utility of exome sequencing as a diagnostic tool.
Exome sequencing identifies mutations in a novel gene in patients with Noonan syndrome. T. Nihon1, Y. Aoki1, T. Banjo2, N. Okamoto3, S. Mizuno1, K. Kuroswaga1, T. Ogata1, F. Takada1, M. Yano1, T. Ando2, T. Hoshika1, C. Barnett1,1,2,1, H. Ohashi1,3, H. Kawarne1,1, T. Hasagawa1,1, T. Okutani1, T. Nagashima1, S. Hasegawa1,3,3, R. Funayama1,1, T. Nagashima1,1, K. Nakayama1,1, S. Inoue1, T. Watanabe2, T. Ogura3, Y. Matsubara1,1, Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Miyagi, Japan; 2) Department of Developmental Neurobiology, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan; 3) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi, Japan; 4) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Kasugai, Japan; 5) Division of Medical Genetics, Kanagawa Children’s Medical Center, Yokohama, Japan; 6) Department of Pediatrics, Hamamatsu University School of Medicine, Hamamatsu, Japan; 7) Department of Medical Genetics, Kitasato University Graduate School of Medical Sciences, Sagamihara, Japan; 8) Department of Pediatrics, Akita University School of Medicine, Akita, Japan; 9) Department of Pediatrics, Municipal Tsuruga Hospital, Tsuruga, Japan; 10) Department of Pediatrics, Tottori Prefectural Central Hospital, Tottori, Japan; 11) South Australian Clinical Genetics Service, SA Pathology, Women's and Children's Hospital, North Adelaide, Australia; 12) School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, South Australia, Australia; 13) Division of Medical Genetics, Saitama Children's Medical Center, Saitama, Japan; 14) Department of Genetic Counseling, Ochanomizu University, Tokyo, Japan; 15) Department of Pediatrics Keio University School of Medicine, Tokyo Japan; 16) Division of NICU, General Perinatal Medical Center, Wakayama Medical University, Wakayama, Japan; 17) Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan; 18) Department of Pediatrics, Niigata Graduate School of Medical and Dental Sciences, Niigata, Japan; 19) Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan.

RAS GTPases mediate a wide variety of cellular functions, including proliferation, cell survival and differentiation. Recent studies have revealed that germline mutations and genetic mosaicism of classical RAS, including KRAS, HRAS and NRAS, are associated with Noonan syndrome and related disorders (RAS/mitogen-activated protein kinase (MAPK) pathway syndromes or RASopathies), nevus sebaceous and Schimelpenning syndrome. In the present study, we identified a total of nine missense, non-synonymous mutations in a gene encoding a member of the RAS family, in 17 of 180 individuals (9%) with Noonan syndrome and related conditions without mutations in known genes. Clinical manifestations in the mutation-positive individuals are consistent with those of Noonan syndrome, which are characterized by distinctive facial appearance, short stature and congenital heart defects. Seventy percent of mutation-positive individuals had hypertrophic cardiomyopathy, a high frequency compared with the 20% incidence in individuals with Noonan syndrome overall. Luciferase assays in NIH3T3 cells showed that five mutants identified in children with Noonan syndrome enhanced ELK1 transactivation. The introduction of mutant mRNAs of this gene into one cell-stage zebrafish embryos was found to result in a significant increase of embryos with craniofacial abnormalities, incomplete looping and a hypoplastic chamber in the heart and an elongated yolk sac. These results demonstrated that gain-of-function mutations in this gene cause Noonan syndrome, showing a similar biological effect to mutations in other RASopathy genes. Note: The name of the gene will be disclosed at the meeting.
3119W
Identification of pathogenic variants in Idiopathic Scoliosis. S. Pat-
ten1,2, E. Aix3,4, A. Labalme1,2, S. Girard6, B. Boit6, C. Poizat1,4, D. Sania-
ville1,4, J. Berard1, G. Rouleau6, F. Clerget-Darpoux1, P. Drapeau1, F. Mol-
dovan2, P. Edery1,4. 1) Pathology and Cell Biology, Université de Montrèal, 
Montreal, Quebec, Canada; 2) CHU Sainte Justine and Faculté de Médecine 
Dentaire, Université de Montréal, Quebec, Canada; 3) Hospices Civils de 
Lyons, Department of Genetics, Lyons, France; 4) INSERM U829, Lyon, 
France; 5) Université Lyon 1, Neuroscience Research Centre, Lyon, France; 
6) Montreal Neurological Institute and Hospital, Department of Neurology 
and Neurosurgery, McGill University, Montreal, Quebec, Canada; 7) Croix-
Rouge French Red Cross Massif Hospit Traumatology, Fontainebleau, 
France. 

Idiopathic scoliosis (IS) is a spine deformity affecting up to 3% of adoles-
cents. Despite strong evidences of genetic contributions to the etiology of 
IS, the causative genes remain unidentified. We previously identified a new 
disease gene location (3q12.1 or 5q13.3) in a large extended family with 

In the present study, we performed SNP genotyping to further refine these 
loci. Whole exome sequencing was performed in 7 affected individuals 
from the large family and 4,564 single nucleotide variants (SNVs) were identified 
in our targeted chromosomal regions. These SNVs were subsequently fil-
tered and the pathogenic nature of candidate SNVs was assessed by func-
tional studies in zebrafish. We identified one rare missense SNV (MAF<1%) 
in a gene, within the 5q13.3 IS interval, that segregated with IS in the family 
and resulted in spinal deformities similar to that observed in IS patients 
when overexpressed in zebrafish. Sanger sequencing of this gene was 
subsequently performed in 40 additional multiplex families, and in 150 unre-
lated IS cases. Co-transmission of that same SNV and the disease was 
also observed in 3/40 additional families and it was identified in 3/150 
unrelated IS cases. We also identified a missense mutation in this gene in 
1/40 family, and another rare missense variant in the gene in 5/150 IS 
patients. None of these three nucleotide variants were found in 206 control 
chromosomes of individuals in the same geographical area. Functional analy-

dies of these mutations in zebrafish also resulted in spine deformities similar 
to that observed in IS patients. These results indicate that these variants 
cause IS and, to our knowledge, we report on the first ever causative gene 
in idiopathic scoliosis.

3120T
Exome sequencing is the preferred approach for identifying the genetic 
cause in consanguinous and non-consanguinous recessive disease. H. Smee-
t1, J. Vanoevelen1, M. Gerards1, R. Kamp3, T. Theunissen1, B. De 
Koning1, I. Boesten1, M. van Geel1, P. Lindsey1, C. Stumpel1, M. Ngu-
yen1, M. Gerrits1, S. Ghesquire1, S. Stevens1, C. de Die1, B. van den 
Bosch1, I. De Coo1, 1) Dept Clin Genet, Maastricht UMC, Maastricht, 
Netherlands; 2) Dept Neurology, Erasmus UMC, Rotterdam, Netherlands.

Exome sequencing is becoming the primary approach in genetic testing, 
particularly suited to solve de novo cases and patients with recessive de-
ease. We performed exome sequencing in 27 families with predominantly 
recessive neurological syndromes and, if available, a metabolic cellular 
phenotype. In most families more than one child was affected. Counselling 
information and SNP-array data was used to classify the parents as consan-
guinous or non-consanguinous. Exome sequencing was done on a single 
patient per family, using an Illumina HiSeq2000 system. Quality criteria were 
>90 million reads per sample, >90% coverage per base at 10X, more than 
65% coverage per base at 20X, <50% duplicate reads, >80X mean corrected 
coverage and a specificity of >65%. Data analysis was performed using an 
in-house optimized pipeline. Filtering of the variants was based on frequency 
(<1%), functional impact, function of the gene, the genetic model (homozig-
gous or compound heterozygous), genetic localization and segregation in 
the family. If the number of remaining variants was too high or the data 
did not match the quality criteria (<10%) an additional patient was 
sequenced. A total of seven patients were homozygous or compound hetero-
zygous for mutations in the genes LPIN1, SCL19A3 (2 times), AARS2, 
DHODH, SERAC1, MMP14, five of which could be directly linked to the 
phenotype and two were confirmed after in vitro validation. In 16 patients 
possibly pathogenic variants were identified in genes, which could be con-

nected to the phenotype. If possible, these variants are functionally validated 
or confirmed in additional patients, but providing the definite evidence can 
be cumbersome. Still, we expect to solve the majority of the families studied. 
In the consanguineous families, the phenotypes are often not caused by a 
single gene defect and more genes are mutated, each explaining different 
parts of the clinical spectrum. As the parents involved have a risks of offspring 
with different pathologies, it is advisable in those families to screen both 
parents completely for shared recessive mutations to determine recurrence 
risk and prenatal options.

3121F
Next-Generation Sequencing Identities PXDN Mutations in Patients 
with Complex Microphthalmia and Anterior Segment Eye Disease. A. 
Slavotinek1, A. Choi2, R. Lao2, P. Ling-Fung Tang2, W. Mayer1, E. Wan2, 
T. Bardakjian3, P. Y. Kwok2, A. Schneider3. 1) Dept Pediatrics, Division of 
Genetics, Univ California, San Francisco, San Francisco, CA; 2) Cardiovas-
cular Research Institute, UCSF, San Francisco, CA; 3) Division of Medical 
Genetics, Einstein Medical Center, Philadelphia, PA.

Anophthalmia and microphthalmia (A/M) are common and signiﬁcant 
because of visual loss. Mutations in SOX2 and other genes can cause A/M, 
but more than half of affected individuals do not receive a molecular 
diagnosis. A/M can be isolated (simplex A/M) or accompanied by additional 
ocular defects (complex A/M). We used exome sequencing to study children 
with simplex A/M (n=14), complex A/M (n=5) and other eye defects including 
anterior segment dysgenesis (ASD) and retinal dysplasia (n=3). Samples 
were initially analyzed for known genes associated with A/M and if negative, 
Further analysis was undertaken. We present two families who had PXDN 
mutations as a cause of their eye defects. The first family had an affected 
sib pair with complex A/M and ASD, sclerocornea, microphthalmia, develop-
mental delays and hypotonus. The older male sibling had glaucoma. Both 
sibs had two PXDN mutations; c.1021C>T, predicting p.Arg341* and a 
framesshift mutation, c.2375_2397del, predicting p.Tyr791Serfs*65 and pre-
mature protein truncation. The second family was a single affected male 
with bilateral ASD who had healthy parents. There was no known consan-
guinity. This child had a paternally inherited, missense mutation, 
p.Gln316Pro, and a maternally inherited frameshift mutation, 
p.Tyr398Hisfs*40. Mutations in PXDN have been described in 3 families 
with congenital cataracts, microcornea, sclerocornea and glaucoma, but 
the large family and 4,564 single nucleotide variants (SNVs) were identified 
in 7 affected individuals from 

3q12.1 or 5q13.3 disease gene location (3q12.1 or 5q13.3) in a large extended family with 

In the present study, we performed SNP genotyping to further refine these 
loci. Whole exome sequencing was performed in 7 affected individuals 
from the large family and 4,564 single nucleotide variants (SNVs) were identified 
in our targeted chromosomal regions. These SNVs were subsequently fil-
tered and the pathogenic nature of candidate SNVs was assessed by func-
tional studies in zebrafish. We identified one rare missense SNV (MAF<1%) 
in a gene, within the 5q13.3 IS interval, that segregated with IS in the family 
and resulted in spinal deformities similar to that observed in IS patients 
when overexpressed in zebrafish. Sanger sequencing of this gene was 
subsequently performed in 40 additional multiplex families, and in 150 unre-
lated IS cases. Co-transmission of that same SNV and the disease was 
also observed in 3/40 additional families and it was identified in 3/150 
unrelated IS cases. We also identified a missense mutation in this gene in 
1/40 family, and another rare missense variant in the gene in 5/150 IS 
patients. None of these three nucleotide variants were found in 206 control 
chromosomes of individuals in the same geographical area. Functional analy-

dies of these mutations in zebrafish also resulted in spine deformities similar 
to that observed in IS patients. These results indicate that these variants 
cause IS and, to our knowledge, we report on the first ever causative gene 
in idiopathic scoliosis.
Diagnostic exome sequencing to elucidate the genetic basis of likely recessive disorders in consanguineous families. P. Makrythanasis1, M. Neils1,2, F.R. Santori3, M. Del Rosso2, A. Vannier1, F. Béna2, S. Gimelli2, E. Stathaki2, S. Tertmayr4, A. Megarbane1,6, E. Masi1, M.S. Aghani3, M.S. Zakr1, A. Bottani2, S. Fokouam1, S. Klitsou-Tzel6, H. Fryssira, E. Kanavakis1, A. Al-Allawi7, A. Setfani, S. Al-Hait12, S.C. Elalouf11, N. Jaakko1, L. Al-Gazali15,14, F. Al-Jasmi3,14, H. Chaabouni Boohermamed15, E. Abitalba16, D.N. Cooper3,17, H. Hamamy7, S.E. Antonarakis1,3,1, 1) Dept Genetic Medicine & Dev, University of Geneva, Geneva, Switzerland; 2) Estonian Genome Centre, University of Tartu, Tartu, Estonia; 3) Service of Genetic Medicine, University Hospitales of Geneva, Geneva, Switzerland; 4) Department of Clinical Genetics, National Research Centre, Cairo, Egypt; 5) Medical Genetics Unit, Saint Joseph University, Beirut, Lebanon; 6) Institut Jérôme Lejeune, Paris, France; 7) Pediatric Department, The University of Jordan, Amman, Jordan; 8) Department of Medical Genetics, University of Athens, Athens, Greece; 9) Department of Pathology, College of Medicine, University of Dohuk, Dohuk, Iraq; 10) Département de Génétique Médicale, Institut National d’Hygiène, Rabat, Morocco; 11) Centre de Génomique Humaine, Faculté de Médecine et de Pharmacologie, Université Mohamed V Souissi, Rabat, Morocco; 12) Genetics & IVF Department, The Farah Hospital, Amman, Jordan; 13) Department of Paediatrics, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates; 14) Department of Pediatrics, Tawam Hospital, United Arab Emirates University, Al Ain, United Arab Emirates; 15) Department of Medical Genetics, University Tunis El Manar, Faculty of Medicine, Tunis, Tunisia; 16) Medical Research Institute, Alexandria University, Alexandria, Egypt; 17) Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, Wales, UK.

Consanguinity is a risk factor for autosomal recessive (AR) disorders. Although many of the clinical phenotypes presenting in the offspring of consanguineous couples are of unknown etiology, advances in sequencing the protein-coding portion of the human genome (exome) provide an opportunity to arrive at a molecular diagnosis in at least a proportion of unresolved phenotypes, thereby identifying novel candidate genes responsible for AR phenotypes. Samples were collected from 41 consanguineous families characterized by the presence of clinical phenotypes consistent with an inherited disorder. DNA was taken from the patient(s), all unaffected siblings and the parents. All samples were genotyped with a 720K SNP array to define the identical-by-descent chromosomal regions likely to contain the responsible pathogenic variants. Exome sequencing was performed on one affected individual per family. Variants within the identified target areas were called and filtered using bioinformatic tools. The putative pathogenic variant was confirmed the mutation to be homozygous in the three affected children and heterozygous in the unaffected siblings. We identified a deleterious homozygous splice acceptor site mutation in intron 13 of UBE3B (c.1451-2A>G) predicted to result in skipping of exon 14. The UBE3B gene resides in a 591kb region of shared autosyogosity on chromosome 12q24.11, identified by SNP array in the affected siblings and affected parental cousin. We confirmed the mutation to be homozygous in the three affected children and heterozygous in both the parents and two unaffected siblings. The features in this family, while distinct, demonstrate significant clinical overlap with those identified in BPIDS and KOS. Our findings suggest that there is a broader clinical spectrum of features caused by mutations in UBE3B, underscoring the evolving genetic landscape of phenotypically overlapping disorders in the neurocognitive syndromic developmental axis.

X-linked Intellectual Disability (XLID) is a group of genetically highly heterogeneous disorders caused by mutations in genes on the X chromosome. More than 90 XLID genes have been identified, which account for ~50% of XLID disorders. To systematically identify responsible genes for the remaining XLID, we sequenced the X-exome of 55 XLID families (a single affected male from 27 families and two affected males from 28 families) using an Agilent SureSelect system and the HiSeq2000 platform. An average of 1.644 ± 246 variants were identified in each X-exome with a mean coverage of 49x. To enrich for disease-causing mutations, we initially tried to use variant datasets from dbsNP and the male portion of 1000 Genomes as filters. However, these databases present several challenges for an efficient enrichment of disease-causing mutations. dbsNP contains potentially unannotated pathological mutations and available male data from 1000 Genomes contain an excess of ambiguous variants for the X chromosome. To solve these problems, we tested and optimized a novel strategy by sequencing affected sibs in proband families to enrich for shared variants and by eliminating neutral variants using a small cohort of unrelated males. This approach achieved a substantial (98.1%) reduction in the number of variants for follow-up studies and significantly enriched for known XLID genes, including ATRX, HUWE1, MECP2, MED12, MEF2A, and P53BP (p < 0.005, hypergeometric test). Novel candidate XLID genes including ZC4H2 are also identified. We conclude that the affected sib-based approach can be used to effectively enrich for disease-causing mutations in genetically heterogeneous X-linked disorders and that publicly available human reference databases including dbsNP and 1000 Genomes should be used with caution as automatic filters to enrich for disease-causing mutations in X-linked disorders.


Purpose: Perrault syndrome, first reported in 1951, is characterized by sensorineural hearing loss in both genders and gonadal dysfunction. Since 1951, 17 cases of Perrault syndrome have been described. Clinical presentations include sensorineural hearing loss, posterior calvarial defects, and uterine anomalies. To systematically identify responsible genes for the remaining XLID, we sequenced the X-exome of 55 XLID families (a single affected male from 27 families and two affected males from 28 families) using an Agilent SureSelect system and the HiSeq2000 platform. An average of 1.644 ± 246 variants were identified in each X-exome with a mean coverage of 49x. To enrich for disease-causing mutations, we initially tried to use variant datasets from dbsNP and the male portion of 1000 Genomes as filters. However, these databases present several challenges for an efficient enrichment of disease-causing mutations. dbsNP contains potentially unannotated pathological mutations and available male data from 1000 Genomes contain an excess of ambiguous variants for the X chromosome. To solve these problems, we tested and optimized a novel strategy by sequencing affected sibs in proband families to enrich for shared variants and by eliminating neutral variants using a small cohort of unrelated males. This approach achieved a substantial (98.1%) reduction in the number of variants for follow-up studies and significantly enriched for known XLID genes, including ATRX, HUWE1, MECP2, MED12, MEF2A, and P53BP (p < 0.005, hypergeometric test). Novel candidate XLID genes including ZC4H2 are also identified. We conclude that the affected sib-based approach can be used to effectively enrich for disease-causing mutations in genetically heterogeneous X-linked disorders and that publicly available human reference databases including dbsNP and 1000 Genomes should be used with caution as automatic filters to enrich for disease-causing mutations in X-linked disorders.


The mitochondrial protein synthesis apparatus allows the synthesis of the 13 respiratory chain (RC) subunits encoded by mitochondrial DNA. The deficiency of mitochondrial protein synthesis can be caused by mutations in any component of the translation apparatus including tRNA, tRNA, and proteins. We report here 4 patients from 3 independent families with mutations in 3 nuclear genes encoding proteins of the mitochondrial translation apparatus. All patients presented a multiple RC deficiency in muscle, liver or fibroblasts and an abnormal pattern assembly of RC complexes in fibroblasts. Interestingly the clinical presentations of the patients differ for the previously reported patients with mutations in these genes. Patients 1-2 are two brothers born to non consanguineous parents. They presented hypoglycemia, truncus hypotonia and lactic acidosis at 2 days of age. They died at 6 and 5 days of age. Exome sequencing detected two heterozygous mutations of MTO1 encoding the mitochondrial translation optimization 1. MTO1 mutations have been identified in two families with hypertrophic cardiomyopathy and lactic acidosis. Patient 3 is a girl born to consanguineous parents of Turkish origin. She has sideroblastic anemia, psychomotor retardation, microcephaly, exocrine pancreatic dysfunction, short stature, and mild mental retardation. She has a homozygous PUS1 mutation. PUS1 encodes the pseudouridylylate synthase 1 that converts uridine to pseudouridine once it has been incorporated into an RNA molecule. PUS1 mutations were previously reported in patients with multiple RC deficiency and myopathy, lactic acidosis, and sideroblastic anemia (MLASA). Patient 4 is a girl born to non consanguineous parents. She presented truncal hypotonia in the first months of life. She developed tonic spasms without hypothermia at the age of 2 years. EEG showed a general slowing in the delta and theta band. MRT detected bilateral and generalized hypotonia of the limbs. Exome sequencing did not identify pathological mutations in GFM1 encoding the mitochondrial elongation factor G1. Until now GFM1 mutations were associated with either severe encephalopathy or liver dysfunction. These results give support to the clinical and genetic heterogeneity of mitochondrial diseases. dbSNP contains potentially unannotated pathological mutations and available male data from 1000 Genomes contain an excess of ambiguous variants for the X chromosome. To solve these problems, we tested and optimized a novel strategy by sequencing affected sibs in proband families to enrich for shared variants and by eliminating neutral variants using a small cohort of unrelated males. This approach achieved a substantial (98.1%) reduction in the number of variants for follow-up studies and significantly enriched for known XLID genes, including ATRX, HUWE1, MECP2, MED12, MEF2A, and P53BP (p < 0.005, hypergeometric test). Novel candidate XLID genes including ZC4H2 are also identified. We conclude that the affected sib-based approach can be used to effectively enrich for disease-causing mutations in genetically heterogeneous X-linked disorders and that publicly available human reference databases including dbsNP and 1000 Genomes should be used with caution as automatic filters to enrich for disease-causing mutations in X-linked disorders.

Targeted sequencing of mitochondrial exome in pediatric patients with mitochondrial diseases. M. Terasova, T. Hanzik, H. Kratuchvila, A. Vondrackova, V. Stranecky, M. Rodinova, H. Hansikova, J. Zeman. 1) Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague 2, Czech Republic; 2) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague 2, Czech Republic.

Mitochondrial diseases (MD) represent both clinically and genetically heterogeneous group of disorders. More than 130 nuclear genes have been described so far whose mutations lead to MD. Accurate targeting of the genetic analysis based on clinical symptoms or laboratory tests is possible for only a few types of MD (e.g. MNGIE syndrome, mitochondrial encephalomyopathy, lactic acidosis and neuropathy - MELAS, Leigh syndrome due to SURF1 mutations). For most of MDs, even specialized enzymatic and protein analyses do not allow unambiguously to narrow a group of candidate genes. Next generation sequencing is a significant milestone in discovery of genetic bases of inherited diseases and could therefore transform genetic diagnostics. In last two years, more than 25 MD genes were identified by exome sequencing. In a group of 28 patients with mitochondrial disease we performed targeted sequencing of mitochondrial exome (>11,000 nuclear genes mostly based on MitoCarta Inventory; http://www.broadinstitute.org/pubs/MitoCarta/index.html). In 6 patients, mutations in known MD-genes (e.g. SURF1, COX10, AIFM1, TK2, and MNGIE) were found. In 3 patients no gene was prioritized. In the other patients candidate variants in genes with previously unknown association with MD were selected and are further evaluated to support their pathogenicity. Supported by RCN V16-0076, V16-0078, RVO 68081723, grants IGA NT14156/3, and grants IGA NT13114/4, IGA NT14156/3.
3128W

Whole Genome Sequencing for Rapid Identification of Sequence Variants Associated with Recessive Canine Disease Models. G.S. Johnson1, D.P. O’Brien2, R. Zeng3, D. Gilliam1, T. Mlhanga-Mutangadura1, J.R. Coates2, M.L. Katz2, J.F. Taylor4, R.D. Schnabel5, 1) Department of Veterinary Pathobiology, University of Missouri, Columbia, MO, USA; 2) Department of Veterinary Medicine and Surgery, University of Missouri, Columbia, MO, USA; 3) Wisconsin Eye Institute, University of Missouri, Columbia, MO, USA; 4) Division of Animal Sciences, University of Missouri, Columbia, MO, USA.

Over the past 25 years we have accumulated over 100,000 canine DNA samples, many from dogs with a variety of inherited diseases. Our main focus is the identification of the causal mutations underlying these diseases with two objectives: (1) to facilitate the breeding of healthy dogs, and (2) to identify canine models for heritable human diseases. In 2010, we began sequencing the genomes of dogs believed to bear distinct recessive diseases. To date, we have generated genome sequence data from 28 dogs. Collaborators at UCLA and T-GEN have provided an additional 34 genome sequences from domestic dogs and closely related wild canids. Ongoing sequencing efforts and collaborations are expected to produce in excess of 100 genome sequences by the end of 2013. Our model for the identification of causal variants uses 1 case and N-1 controls as a starting point, where the case is assumed to be homozygous for a variant that is not present in any of the N-1 control animals. Additional models under development will accommodate more complex genetic etiologies such as compound heterozygosity and incomplete penetrance. Currently, our efforts have resulted in the identification of five variants that are apparently causal for their respective diseases. We shall describe an SAG mutation associated with progressive retinal atrophy as an example. Furthermore, these sequencing, genome-specific assays testing for known variants and Sanger sequencing of specific gene regions provide answers for some patients. We applied whole exome sequencing in a family with LCA, where previous targeted assay and sequencing approaches had been unsuccessful in disease gene identification. We implemented two independent bioinformatic pipelines, BWA-GATK and SOAP, to interrogate the sequence data and maximize our capacity for valid variant identification. Our first pipeline resulted in 16 prioritized variants, while the second led to 13. Four of these variants, two groups of prioritized variants, one of which was the top candidate predicted from both pipelines, a novel homozygous nonsense mutation in TULP1. TULP1 is a member of the tubby gene family and has roles in rhodopsin movement, synapse formation and survival of the photoreceptors, and causes approximately 1% to 2% of cases with LCA. The disease-causing variant was confirmed on Sanger sequencing and segregated appropriately in the family. This study highlights the value of a whole exome sequencing approach as a rapid means for known and novel variant identification in disease genes in the genetically heterogeneous retinal disease of LCA.

3129T

Whole Exome Sequencing on two fetuses with Centronuclear Myopathy of consanguineous parents from Sudan shows two possible candidates. M.C. Coates1, A. Petersen2, P. Skovbo2, I.S. Pedersen3, A. Ernst2, H. Krarup2, M.B. Petersen1, 1) Department of Clinical Genetics, Aalborg University Hospital, Denmark; 2) Section of Molecular Diagnostics at Department of Clinical Biochemistry, Aalborg University Hospital, Denmark; 3) Institute of Pathology, Aalborg University Hospital, Denmark; 4) Department of Obstetrics and Gynaecology, Aalborg University Hospital, Denmark.

Arthrogryposis multiplex congenita (AMC) is a rare congenital disorder characterized by multiple joint contractures. There are several possible underlying conditions but the common pathway is the lack of fetal movement. Usually termination of pregnancy is performed due to the known severity of the disorder. Exact diagnosis is identified at autopsy subsequently. We present a consanguineous healthy couple from Wadi Hafa, Sudan, with two separate pregnancies with female fetuses showing AMC. The mother’s great grandfather was the grandfather of the father. Ultra sound (US) in the 13th week showed AMC, leading to termination of the first pregnancy. The second pregnancy showed the same malformation on US in the 13th week, also leading to termination. Array-CGH, karyotype (46,XX) and investigation for Congenital Myotonic Dystrophy were all normal. A full autopsy including CNS and muscle histopathological examination of the two foetuses were performed. The muscle examination included immunohistochemistry and electron microscopy. A diagnosis of Centronuclear Myopathy with autosomal recessive inheritance (AR-CNMy) was proposed as there was a greater amount of centrally located nuclei and a greater variation in the muscle fibre diameter than expected for the gestational age and as no other specific myopathies were found. CNM is seen with either X-linked, autosomal dominant or recessive inheritance. The autosomal recessive form of CNM is currently only reported to be caused by mutations in either the BIN1 or RYR1 genes, and many patients are without any genetic clarification. Some cases were detected during fetal life, but none of them were severe enough to result in termination of pregnancy, in contrast to our case. Blood samples and cell cultures from the parents and foetuses, respectively, were analyzed with Whole Exome Sequencing (WES). The WES result showed no mutations in the two previously known genes, RYR1 and BIN1. This is in agreement with results from sanger sequencing previously carried out on BIN1. The WES suggested two possible candidates for a novel gene responsible for AR-CNMy. They were identified using multiple filtration parameters to filter from the starting collections of twelve thousands variants for each of the four exomes. Both genes are without any previous connection to a neuromuscular disease. We are in the process of carrying out confirmation on our WES results concerning the two candidates.

3130F

Value of whole exome sequencing for novel variant identification in Leber congenital amaurosis. Y. Guo1,2, J. Prokudin1,2,3,4, C. Yu1,2,4, J. Liang1, Y. Xie1, M. Fisherty2, L. Tian2, S. Crofts3, F. Wang2, J. Snyder2, C. Donaldson2, N. Abdel-Magid1, L. Vazquez1, B. Keating2,7,7, H. Hakonarson1,7,8,9, J. Wang5,6,10,11,15, R. Jamieson2,3,12,13,15, 1) Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, PA, USA; 2) Eye and Developmental Genetics Research Group, Western Sydney Genetics Program, The Children’s Hospital at Westmead, Sydney, NSW, Australia; 3) Children’s Medical Research Institute, Westmead, Sydney, NSW, Australia; 4) College of Life Sciences, Sichuan University, Key laboratory for Bio-resources and Eco-environment of Ministry of Education, Sichuan Key Laboratory of Molecular Biology and Biotechnology, Chengdu 610064, PR China; 5) BGI-Shenzhen, Shenzhen 518083, China; 6) Department of Ophthalmology, The Children’s Hospital at Westmead, Sydney, NSW, Australia; 7) Division of Human Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA, USA; 8) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 9) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 10) King Abdulaziz University, Jeddah, Saudi Arabia; 11) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark; 12) Discipline of Ophthalmology & Save Sight Institute, University of Sydney, Sydney, Australia; 13) Disciplines of Paediatrics and Child Health & Genetic Medicine, University of Sydney, Sydney, NSW, Australia; 14) Equal contribution; 15) Corresponding authors.

Leber congenital amaurosis (LCA) is a severe form of retinal dystrophy and children present with nystagmus and poor vision in the first year of life. There are at least 21 known disease genes and few distinguishing clinical examples. Furthermore, these sequencing, genome specific assays testing for known variants and Sanger sequencing of specific gene regions provide answers for some patients. We applied whole exome sequencing in a family with LCA, where previous targeted assay and sequencing approaches had been unsuccessful in disease gene identification. We implemented two independent bioinformatic pipelines, BWA-GATK and SOAP, to interrogate the sequence data and maximize our capacity for valid variant identification. Our first pipeline resulted in 16 prioritized variants, while the second led to 13. Four of these variants, two groups of prioritized variants, one of which was the top candidate predicted from both pipelines, a novel homozygous nonsense mutation in TULP1. TULP1 is a member of the tubby gene family and has roles in rhodopsin movement, synapse formation and survival of the photoreceptors, and causes approximately 1% to 2% of cases with LCA. The disease-causing variant was confirmed on Sanger sequencing and segregated appropriately in the family. This study highlights the value of a whole exome sequencing approach as a rapid means for known and novel variant identification in disease genes in the genetically heterogeneous retinal disease of LCA.
**3131W**


1) Equipe d’accueil EA 4271 GAD “Génétique des Anomalies du Développement”, IFR Santé STIC, Université de Bourgogne, Dijon, France; 2) Centre de Référence Anomalies de Développement et Syndromes Malformatifs de l’Interrégion Grand-Est, Hôpital d’Enfants, CHU, Dijon, France; 3) Centre de Génétique, Hôpital Enfants, CHU, Dijon, France; 4) Département de Génétique, Hôpital Necker-Enfants Malades, APHP, Paris, France; 5) Département d’Anatomopathologie, Centre de Biologie et de Pathologie, CHRU, Lille, France; 6) Service de Génétique Médicale et Centre de Référence Anomalies du Développement et Syndromes Malformatifs, CHU de Bordeaux,Laboratoire MRGM, EA4576, Université de Bordeaux, Bordeaux, France; 8) Département de génétique, Laboratoire Cerba - Saint-Ouen l’Aumone et Service de Gynécologie-Obstétrique, Hôpital R. Dubos, Pontoise France; 9) Department of Pediatrics, Shimada Center for Rare Diseases and Neurodevelopmental Disorders, Tokyo, Japan; 10) Department of Pediatrics, Kameda Medical Center, Kamao-gawa, Japan; 11) Unité INSERM U781 et Fondation IMAGINE, Hôpital Necker-Enfants-Malades, Paris, France; 12) AP-HP, Unité de Fetoopathologie, Hôpital Antoine Béclère, Clamart, France; 13) Département de Fetoopathologie et Anatomopathologie, Laboratoire Cerba - Saint-Ouen l’Aumone, France; 14) Service de Gynécologie-Obstétrique, Hôpital R. Dubos, Pontoise, France; 15) Laboratory of Human Embryology, Institute of Medical Biology, Agency for Science, Technology and Research (A*STAR), Singapore; 16) Farah University, Faculty of Biomedical Science, Iran; 17) Laboratoire de Génétique Moléculaire, Pla-teau Technique de Biologie, CHU Dijon, France; 21) Service de Pédiatrie 1, Hôpital d’Enfants, CHU, Dijon, France; 22) Medical Genetics Department, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey; 23) Sagop Unité de Génétique Médicale, CHU, Nantes, France; 24) Téléthon Institute of Genetics and Medicine, Napoli, Italy, Medical Genetics, Department of Medical Translational Sciences, University of Naples Federico II, Napoli, Italy.

The oral-facial-digital syndrome type VI (ODF VI) or Varadi-Papp syndrome has recently been characterised by the following diagnostic criteria: molar tooth sign (MTS) and one or more of the following: 1) tongue hamartoma and/or additional frenula and/or upper lip notch; 2) mesoaxial polydactyly of one or more hands or feet and 3) hypothyalmic hamartoma. Because of the MTS, ODF VI belongs to the ‘Joubert Syndrome Related Disorders’ (JSRD). Its genetic etiology remains largely unknown although mutations in the TMEM216 gene, responsible for Joubert (JBS2) and Meckel-Gruber (MKD) syndromes, were identified in two patient families. In total, 20 patients were identified in the X-linked ODF1 gene have also been reported in male JBS patients with oral defects or female patients with ODF syndrome and male MTs. Combined exon and Sanger sequencing identified compound heterozygous or homozygous causal mutations in the C5orf42 gene in 8/9 families including a severe fetus with microphthalmia, cerebellar vermis hypoplasia, corpus callosum agenesis, bilateral hand and feet preaxial and postaxial polydactyly, unilateral hand mesoaxial polydactyly with Y-shaped metacarpal phalange, and severe skeletal dysplasia but absent oral manifestations. Altogether, we identified 12 novel C5orf42 mutations in 11 cases from 8 families. All patients presented distal anomalies with constant feet preaxial polydactyly (11/11 cases) and frequent mesoaxial polydactyly (7/11 cases) in at least one distal extremity with Y-shaped metacarpal abnormality (8/8 cases). Mutations of this gene have already been reported in Joubert but classical additional features usually reported in JSRD appeared absent, such as poly cystic disease and retinal disease. In contrast to highly heterogeneous JSRD with 20 causative genes identified to date, ODF VI seemed genetically highly homogeneous. In conclusion, we report that C5orf42 is the major gene responsible for ODF VI syndrome and confirm that ODF VI and JBS syndromes are allelic disorders, enhancing the complexity, highly heterogeneous group of ciliopathies.

**3132T**


1) Brigham and Women’s Hospital, Harvard Medical School, Boston, MA; 2) Uniformed Services University of the Health Sciences, Bethesda, MD; 3) Cardiovascular and Pulmonary Branch, NHLBI, National Institutes of Health, Bethesda, MD; 4) National Tuberous Sclerosis Complex (TSC) is an autosomal dominant disorder caused by mutations in TSC1 or TSC2. It is characterized by formation of tumors that are thought to arise through a two-hit gene inactivation model. Monogenetic TSC cases have been described, and mosaicism is known to occur at low frequency in TSC. We sought to examine several questions on the pathogenesis of TSC skin lesions through analysis of a set of TSC patients enriched for those with adult presentation and diagnosis (PMID 21690595).

From 18 TSC2 patients, 23 TSC patients and 23 parallel samples of normal appearing skin. Long range PCR was used to amplify all coding regions and most nearly intrinsic sequence from each of TSC1 and TSC2, followed by library preparation for next generation sequencing. Over 3000x read depth was achieved, and sequencing data was analyzed using BWA, Picard, GATK, Python, and Matlab to identify all sequence variants present at an allele frequency ≥1%. We identified TSC2 mutations in 27 of 33 cultures, with allele frequencies ranging from 1% to 56%, and a single TSC1 mutation in one culture. Five cultures, from patients with adult presentation of TSC, had no mutations in either TSC1 or TSC2. Sixteen cultures had two different mutations in TSC2, including one case of loss of heterozygosity (LOH). Twelve had only a single mutation identified. Seven (39%) of 18 patients showed mosaicism. Mosaicism was defined as the presence of at least one somatic TSC2 mutation seen in the cultured cells. In four patients, multiple tumors showed distinct second hit mutations in TSC2. Strikingly, 7 of the 14 second hit point mutations were CC > TT dinucleotide mutations, likely due to UV-damaged DNA repair. A germline mosaic TSC2 mutation was identified in four patients and a germ line mosaic TSC2 mutation in TSC1 or TSC2. In summary, there are multiple novel observations: 1) Skin tumors in TSC develop through two-hit inactivation of TSC2. 2) Different TSC skin lesions have distinct second hit events, indicating that they must arise independently. 3) Genomic or classic LOH is rare in TSC skin lesions. 4) UV-induced mutation is likely a contributing genetic event to development of skin lesions in TSC. 5) Mosaicism is common in TSC patients with adult presentation. 6) Lack of mutation identification in skin tumors from TSC patients with adult presentation is consistent with extreme mosaicism as a potential cause.

**3133F**

REPS1 is a novel gene of Neurodegeneration with Brain Iron Accumulation (NBIA). A. DRECOUD1, N. BODDAERT2, I. DESGUERRE3, D. CHRETIEN1,2, A. MUNNICH1,2, A. ROTIG1, I. INSTITUT INSERM, INSERM U781, PARIS, France; 2) DEPARTEMENT DE PEDIATRIE, HOPITAL NECKER-ENFANTS MALADES, PARIS, FRANCE.

Neurodegeneration with brain iron accumulation (NBIA) encompasses a group of rare neurodegenerative disorders transmitted with an autosomal recessive inheritance. This heterogeneous group of disorders can be differentiated by clinical, brain MRI and molecular features. The hallmark feature of NBIA includes progressive neurological deterioration (dystonia, rigidity, choreothetosis) and iron accumulation in the brain, primarily in the basal ganglia. To date, mutations in PANK2, PLA2G6, FA2H, ATP1A3, C20orf37, CP, FTL and WDR45 have been associated with NBIA. The largest subgroup of NBIA cases is caused by mutations in PANK2. The function of REPS1 in iron metabolism is currently unknown but it has been well demonstrated that its protein interacting partner, RabB11/FIP2, functions in transferrin recycling. Western Blot analysis detected a low Reb1 protein level in patient fibroblasts indicating a probable misfolding and degradation of the protein due to the mutation. The REPS1 transcript was found in normal amount. We investigated the iron metabolism and oxidative stress in patient fibroblasts. Colorimetric ferrozine-based assay allowed us to detect a dramatic increase of iron levels in patient fibroblasts. We also observed an increase of ferritin and iron responsive protein (IRP1) amounts suggesting iron accumulation in those cells. In keeping with that, SOD2 and IRP1 protein amounts were highly increased whereas aconitase activity was decreased. All these results indicate that mutations in REPS1 are likely to induce deregulation of iron metabolism and oxidative stress. In conclusion we report REPS1 as a novel gene for NBIA. Improvement in our understanding of the biochemistry and pathophysiology of this form of NBIA will help for novel therapeutics of this neurological condition.
3134W
Challenging diagnostic cases resolved by whole exome sequencing. S.F. Suchy1, J.E. Abdennur2, T. Ben-Onn3, W.K. Chung4, K. Leydiker4, S.J. Balle1, A. Daly1, E.V. Haverfield1, 1) GeneDx, Gaithersburg, MD; 2) Division of Metabolic Disorders, CHOC Children's, Orange CA; 3) Section of Clinical and Metabolic Genetics, Department of Pediatrics, Hamad Medical Corporation, Weill Cornell Medical College, Doha, Qatar; 4) Columbia University Medical Center, New York, NY.

Accurate diagnosis is essential to the appropriate management of genetic disorders, particularly inborn errors of metabolism (IEM). We present three challenging cases in which the molecular cause was identified by whole exome sequencing (WES). Case 1: A one year-old child presented with hypotonia, developmental delay, encephalopathy, seizures, intermittent strabismus and dystonia. There were two similarly affected siblings. Extensive work-up of blood and urine, and imaging studies were non-revealing. CSF studies were normal. Further genetic testing of the parents. WES was performed (Agilent SureSelect XT2 All Exon V4 kit and an Illumina HiSeq2000), on the proband and parents. A homozygous p.Phe375Leu mutation in the tyrosine hydroxylase gene (TH) was identified. Both parents were carriers of the mutation. Case 2: A two year-old male presented with congenital ichthyosis, developmental delay, while matter changes, mild hip dysplasia and constipation. Studies included: a normal metabolic work-up (organic acids, very long chain fatty acids, amino acids, acylcarnitine profile) and a normal chromosome microarray. WES revealed two mutations in trans in the ALDH3A2 gene. One was a frameshift (p.Asn255IfelsX11) and the second was a deletion of exons 1-3. Case 3: A 3 1/2 year-old male was referred from neurology to the metabolic genetics clinic due to distal muscle weakness, motor delay, an unbalanced gait and speech delay. He had no cleft hypertrophy or scoliosis. The metabolic work-up was performed in-kind. Extensive work-up and parental testing were normal. A novel truncating mutation in the ASXL3 gene was identified. Whole exome sequencing generated from a patient with an atypical presentation can further confound the diagnosis. WES offers another option for these diagnostic dilemmas.

3135T
Exome Sequencing Yield for Implication of Genes in Mendelian Disorders. S. Jiang1, J. Bainbridge1, C. Gonzaga-Jauregui1, W. Wiszniowski1, T. Gambin2, J. Lu1, H. Daddapani1, M. Wang1, C. Buhay1, H. Dinh3, Y. Han1, C. Koval1, J. Santibanez1, J. Reid1, D. Valle1, E. Boerwinkle1, D. Muzny1, J. Lupski1, R. Gibbs1, Centers for Mendelian Genomics. 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 3) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA.

Whole genome sequencing has been successful way to elucidate the molecular basis of Mendelian disorders. The Human Genome Sequencing Center at Baylor College of Medicine has generated over 25 Tb of data for more than 250 phenotypes. Our high-throughput multiplexed exome pipeline, producing 2K capture libraries per month, is automated from library construction through annotated VCF generation in part using the Mercury pipeline. Since 2012, we joined the Centers for Mendelian Genomics with the Baylor-Johns Hopkins Center for Mendelian Genomics with the Baylor-Johns Hopkins Center for Mendelian Genomics with 86 institutions. The collection encompasses over 200 phenotypes and has generated 20 Tb of whole exome Illumina data. Coverage of the genome’s most understood regions is far greater with exome sequencing and still the most cost efficient method for gene association to disease. Whole genome shotgun sequencing generated from a patient with Charcot Marie Tooth syndrome was compared against whole exome data from the same individual. Not only were the majority of disease associated mutations found in the exome data, including the causative CMTM1 focus in the pancreas, but the additional mutations found in SH2T2 because of increased target coverage using this application.

Gene-disease association is still highly debated; however, discovery of a candidate mutation in several unrelated individuals with similar phenotypes provides strong evidence for pathogenicity. We present here a novel pathogenic mutation identified in one family with a novel phenotype, sharing characteristics to Bohring-Opitz syndrome. The proband showed a novo truncating mutation in ASXL3. Additional patients were found searching for other patients with the same disease. This was confirmation of the hypothesis that molecular diagnostics is the future of the whole. Whole exome data in conjunction with the molecular diagnostics will play an integral role in disease associations to genes.

3136F
Autosomal Dominant FSGS without extra-renal symptoms due to mutations in LMX1B gene. K.J. Johnson1, O. Boyer1,2,3, S. Woerner,4 F. Yang, E.O. Oakley5, B. Linghu1, O. Gribouval6, M.J. Tete2,7,8, J.S. Duca9,10, L. Klickstein1, A.J. Damask1, J.D. Szustakowski,1, F. Heibl,9, M. Matignon9, V. Baudouin9, F. Chantrel10, J. Champigneulle11, L. Martin12, P. Nitschke13 M-C. Gubler13, S-D. Chiboust13, C. Antignac13, 1) Novartis Institutes for Biomedical Research, 2) Department of Pediatrics, 3) Department of Hematology, 4) University Hospital of Strasbourg, Strasbourg, France; 5) Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris; 6) Novartis Institutes for Biomedical Research, Basel, Switzerland; 7) AURAL and Service de Néphrologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France; 8) Service de Biochimie, Hôpital Neurologique, Hôpital de la Médicocine Paris XII, Assistance Publique-Hôpitaux de Paris, Paris, France; 9) Service de Néphrologie Pédiatrique, Hôpital Robert Debré, MARHEA, Université Denis Diderot, Assistance Publique-Hôpitaux de Paris, Paris, France; 10) Service de Néphrologie et Médecine Interne, Centre Hospitalier de Mulhouse, Mulhouse, France; 11) Service de Pathologie, CHU Nancy Brabois, Vandoeuvre, France; 12) Service de Pathologie, et Inserm UMR1098, CHU Dijon, Dijon, France; 13) Plateforme de Bioinformatique, Hôpital Necker, Enfants Malades, Paris, France.

Inherited forms of Focal segmental glomerular sclerosis (FSGS) are genetically heterogeneous and mutations in six genes (WT1, LMX1B, ACTN4, TRPC6, INF2 and ARHGAP2) have been reported in about 20% of families with autosomal dominant (AD) forms of the disease. In some AD cases mutations have been identified in the ARHGAP2 gene. A candidate mutation in several unrelated individuals with similar phenotypes was discovered identified a mutation in LMX1B that segregated with FSGS in 5 affected individuals, all of whom lack the extra-renal features of FSGS, including the GBM ultrastructural changes. Screening of the 73 other families in the cohort identified mutations in the same codon (R246) in two further families. Protein structure modeling suggests that R246 is a critical residue for interaction of LMX1B with the homeodomain of target genes. The results indicate that isolated FSGS can arise due to mutations in LMX1B, a gene normally associated with a broader syndromic phenotype. Other genes that underlie syndromes of which FSGS is a part of the phenotype should be considered in screening families in which no underlying cause of their familial FSGS has been identified.
Leber congenital amaurosis with early-onset severe macular atrophy and optic atrophy is likely pathognomonic of NMNAT1 mutations. I. Perrault1, M. Koulouz1, S. Hanein2, N. Delphin3, B. Gilbert-Dussardier4, C. Vincent-Delorme5, C. Edelson6, C. Hamel5, E. Silva5, S. Defoort-Delhemes6, L. Fares-Taie1, S. Gerber7, X. Gerard7, A. Goldenberg8, A. Duncome9, G. Le Meur10, P. Calvas10, A. Munnich10, O. Roche10, H. Dollfus10, J. Kaplan1, J. Rozet1. 1) Genetics of retinal degeneration, INSERM U781, IMAGINE, Paris, France; 2) Centre Hospitalier Universitaire de Poitiers-Poitiers, FRANCE; 3) Centre Hospitalier d’Arras-Arras, FRANCE; 4) Department of Ophthalmology, Fondation Ophtalmologique Adolphe de Rothschild, Paris, France; 5) Neurosciences Institute, Hospital Saint Eloi, Montpellier, France; 6) Department of Ophthalmology, Coimbra University Hospital, Coimbra, Portugal; 7) Department of Exploration of Vision and Neuro-Ophthalmology, Hôpital Roger Salengro, Centre Hospitalier Universitaire Régional, Lille, France; 8) Department of Genetics, Centre Hospitalier Universitaire, Rouen, France; 9) Department of Ophthalmology, Centre Hospitalier Universitaire, Nantes, France; 10) Department of Medical Genetics, Purpan Hospital, Centre Hospitalier Universitaire, Toulouse, France; 11) Department of Ophthalmology, Université Paris Descartes-Sorbonne Paris Cité, Paris, France; 12) Department of Clinical Genetics, Strasbourg University Hospital, Strasbourg, France.

Introduction: The nuclear nicotinamide mononucleotide adenyltransferase 1 (NMNAT1) encodes a homohexameric NAD-synthesizing enzyme as well as a chaperone that protects against neuronal activity-induced degeneration. Recently, NMNAT1 mutations have been reported to cause a highly specific Leber congenital amaurosis phenotype characterized by severe neonatal neurodegeneration of the central retina with early-onset optic atrophy. The purpose of the present study was to search for second NMNAT1 disease alleles in single heterozygote patients harboring the NMNAT1 phenotype.

Methods: Nine sporadic cases and two sibs harboring single heterozygote NMNAT1 mutations were screened for copy number variations (CNV) using array comparative genomic hybridization (CGH, CytoScanHD) and mutations affecting regulatory elements or the splicing using Sanger sequencing.

Conclusion: Here, we report that at least 6/10 LCA patients with single heterozygote NMNAT1 mutations carry a second disease allele undetected by exome sequencing. This result suggests that severe neonatal neurodegeneration of the central retina with early-onset optic atrophy is pathognomonic of NMNAT1 mutations. In-depth molecular analysis of the gene and surrounding regulatory elements should be considered in all patients harboring this highly specific LCA phenotype.
Familial generalized seizures due to LGI1 mutation: importance of family history for genetic testing. D.R.M. Amrani1, J.I. F. Andenmann2, E. Andenmann3, C.L. Raggio4, 5, 6, 7, 8, J. Davalos. Area of Health # 1, Quito, Pichincha, Ecuador.

Case Report A 44-year-old woman, using by 13 years a IUD, asymptomatic, the IUD was removed and we found a 18 centimeters pelvic mass and a red spotty skin pigmentation in the left arm. The ultrasound scan reported a solid mass with heterogeneous echogenicity, with clear margins. The computed tomography scan revealed a mesenteric cyst, the mass measured 14.1 x 10.8 cm and contained approximately one liter of serous fluid and also reported a small hepatic cyst. Discussion Mesenteric cysts are one of the most intra-abdominal tumors. 1 The reported incidence ranges from 1/2,000 to 1/2,500,000 admissions to hospital.2 As proposed by Gross, mesenteric and omental cysts are thought to represent benign proliferations of ectopic lymphatics that lack communication with the normal lymphatic system. Other etiologic theories include (1) failure of the embryonic lymph channels to join the venous system, (2) failure of the leaves of the mesentery to fuse, (3) trauma, (4) neoplasia, and (5) degeneration of lymph nodes.4 Genetic Considerations Carney complex (CNC) is a familial multiple neoplasia syndrome transmitted as an autosomal dominant trait. CNC was initially described as the association of myxomas, spotty skin pigmentation, and endocrine overactivity. A variety of endocrine and nonendocrine tumors occur in patients with CNC. Primary pigment nodular adenocortical disease (PPNAD), a rare cause of ACTH-dependent Cushing syndrome. PPNAD is observed in one-fourth of patients with CNC. Approximately half of the cases of CNC are familial. Putative genetic loci have been identified by linkage analysis at chromosome 2p16 and 17q22-24. Recently, the responsible gene on 17q22-24, PRKAR1A, was identified. Mutations in the PRKAR1A gene cause familial isolated precocious puberty (FIPPP).5 6, 7 8, 9 10 11 12

Posters: Molecular Basis of Mendelian Disorders

3141T Familial generalized seizures due to LGI1 mutation: importance of family history for genetic testing. D.R.M. Amrani1, J.I. F. Andenmann2, E. Andenmann3, C.L. Raggio4, 5, 6, 7, 8, J. Davalos. Area of Health # 1, Quito, Pichincha, Ecuador.

3142F Type V Osteogenesis Imperfecta: don’t miss the subtleties. E.M. Carter1, C.L. Raggio1, J.G. Davis4, 6, 7, 8, 9, 10 11 12 13 14 15

3144T The Role of Molecular Genetic Analysis in the Diagnosis of Primary Ciliary Dyskinesia. R.H. Kim1, D.A. Hall2, E. Cutt3, M.R. Knowles4, K. Nelligan5, K. Nykamp6, M.A. Zarivala7, S.D. Dell8, 9, 10, 11

1) Clinical Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Respiratory Medicine, St Michael’s Hospital, University of Toronto, Toronto, ON, Canada; 3) Division of Pathology, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 4) Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina USA; 5) Division of Genetics, Children’s Hospital Boston, Boston, Massachusetts, USA; 6) Prevention Genetics, Marshfield, Wisconsin, USA; 7) Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina USA; 8) Division of Respiratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 9) Department of Pediatrics and Institute of Health Policy, Management and Education, University of Toronto, Toronto, ON, Canada.

Rationale: Primary Ciliary Dyskinesia (PCD) is an autosomal recessive genetic disorder of motile cilia. The diagnosis of PCD has previously relied on ciliary analysis with transmission electron microscopy or video microscopy. However, patients with PCD may have normal ultrastructural appearance and ciliary analysis has limited accessibility. Alternatively, PCD can be diagnosed by demonstrating biaxial mutations in known PCD genes. Genetic testing is emerging as a diagnostic tool to complement ciliary analysis where interpretation and access may delay diagnosis. Objectives: To determine the clinical utility of genetic testing of patients with a confirmed or suspected diagnosis of PCD in a multi-ethnic urban centre. Methods: 28 individuals with confirmed PCD on ciliary analysis were included in this study. DNA was extracted from peripheral blood and subjected to next generation sequencing using a panel of 22 genes associated with PCD. Results: 28/49 (57%) who underwent ciliary diagnosis had a mutation detected. Of the 22 genes, 19/22 (86%) had mutations in PCDP1, PCDP2, and PCDP3. 27/28 (96%) had at least one mutation in a clinically relevant gene. Conclusions: Genetic testing using a panel of 22 genes is feasible, and shows promise in providing a diagnosis in a large proportion of individuals with PCD. However, the clinical utility of genetic testing and how it is integrated into clinical care remains to be determined.
3145F  
**Genotype-Phenotype Correlation in Bardet-Biedl syndrome.**  
O. MHAMDIB, H. CHABBOUNI1,2,1. 1) Department of Human Genetics, Faculty of Medicines, Tunis, Tunisia; 2) Department of Hereditary disorders, Charles Nicolle Hospital. Boulevard 9 Avril 1006 Tunis.  
Bardet-Biedl syndrome (BBS) is an autosomal recessive disease characterized by retinal dystrophy, obesity, postural polydactyly, learning disabilities, renal involvement, endocrine hypogonadism. BBS is genetically heterogeneous with 17 genes (BBS1-BBS17), mutations in BBS genes accounting for approximately 70-80% of cases. Triallelic inheritance has been suggested in about 5% of cases. The genotype-phenotype correlation in BBS was unclear and can be explained by genetic locus heterogeneity, triallelism and modifier genes such as CCDC28B. In order to analyze more clearly, we evaluated a world wide cohort of 371 patients from 330 families with Bardet-Biedl syndrome in whom disease-causing genes were identified. The phenotype spectrum of studied BBS patients was ranged between mild to severe. Thus, our studies have important implication for molecular diagnosis and genetic counselling in BBS patients.

3146W  
**Fragile X triplet repeat instability is influenced by both the presence and position of AGG interruptions in intermediate and premutation alleles.** S. Nolin1, R. Cao2, J. Taylor2, A. Glicksman3, N. Ersalesi1, W.T. Brown1, J. Coppingier2, G. Latham1, A. Hadd1. 1) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY; 2) Asuragen, Inc., Austin, TX.  
Fragile X syndrome is a common cause of inherited intellectual disability that results from the expansion of a CGG repeat in the 5’ untranslated region of the FMR1 gene to >200 copies. We examined FMR1 repeat instability in 762 transmissions of fragile X intermediate (45-54 repeats) and small premutation (55-90 repeats) alleles. Novel PCR assays were used to identify the number and position of the AGGs that occur within the CGG repeat region. Twenty-eight percent of the transmissions contained no AGGs, 44% had one AGG and 27% had two AGGs. Consistent with prior studies, the absence of AGG interruptions within the repeat was associated with an increased risk for instability on transmission as well as an increased risk for expansion to a full mutation in larger alleles. Alleles with two AGGs were least likely to exhibit instability or expansion to a full mutation on transmission. Alleles with one AGG had moderate risks for instability or expansion to full mutation. Surprisingly, the position of the single AGG interruption in these alleles was correlated with instability. Most of the one-AGG alleles had an interruption at position 10 while a minority had one at position 11. Analysis of one-AGG alleles with 45-59 repeats revealed an increased risk of instability for alleles with an AGG at position 10 compared to those with an AGG at position 11. Similarly, analysis of one-AGG alleles with 70-90 repeats revealed an increased risk for full mutation expansion for alleles with an AGG at position 10 compared to those with the AGG at position 11. These results and the fact that AGGs in expanded alleles occur more frequently at position 10 than position 11, suggest that AGGs at position 11 have a greater influence in stabilizing the FMR1 repeat region on transmission.

3147T  
**Identification of a novel nonsense mutation and a missense substitution in the AGPAT2 gene causing congenital generalized lipodystrophy type 1.** H. Haghighi1, M. Razzaghy-Azar1, A. Tala2, M. Sadeghian1, S. Ellard4, A. Haghighi1. 1) University of Toronto, Toronto, Ontario, Canada; 2) Endocrinology and Metabolism Research Center, Tehran University of Medical Sciences, Tehran, Iran; 3) Pediatric Gastroenterology Department, H. Asiasghar Hospital, Tehran University of Medical Sciences, Tehran, Iran; 4) Institute of Biomedical and Clinical Science, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7PS, UK.  
Congenital generalized lipodystrophy (CGL) is an autosomal recessive disease characterized by the generalized scnat of adipose tissue. CGL type 1 is caused by mutations in gene encoding 1-acylglycerol-3-phosphate O-acyltransferase-2 (AGPAT2). A clinical and molecular genetic investigation was performed in affected and unaffected members of two families with CGL type 1. The AGPAT2 coding region was sequenced in index cases of the two families. The presence of the identified mutations in relevant parents was tested. We identified a novel nonsense mutation (c.685G>T, p.Glu229*) and a missense substitution (c.514G>A, p.Glu172Lys). The unaffected parents in both families were heterozygous carrier of the relevant mutation. The results expand genotype-phenotype spectrum in CGL1 and will have implications in prenatal and early diagnosis of the disease. This is the first report of Persian families identified with AGPAT2 mutations.

3148F  
**Phenotype - genotype correlation in a Colombian Oculocutaneous Albinism (OCA) cohort.** C. Lattig1, D. Sanabria1, A. Fernandez2, O. Urtaltz3, Fundacion Contraste - Albinos por Colombia. 1) Laboratorio de Genética Humana, Departamento de Ciencias Biológicas, Universidad de los Andes, Bogota, Colombia; 2) Hospital Militar, Bogotá, Colombia.  
Oculocutaneous albinism (OCA) is a genetic condition of melanin synthesis characterized by a complete lack or generalized reduction in pigmentation of hair, skin and eyes. OCA1, one of the four types of OCA is the most frequent worldwide and is caused by mutations in TYR gene. The TYR gene consists of 5 exons spanning about 65 kb of genomic DNA and encoding a protein of 529 amino acids. In the present study, we have 20 unrelated cases of albinism with complete ophthalmological evaluation. Sequencing of the TYR gene in these individuals revealed three novel mutations, one nonsense mutation, c.551C>G (S184X) and two missense mutation, c.739T>C (C246R) and c.163 T>G (p.C55S), all of them in exon 1. We also found the p.R402Q temperature-sensitive mutation in three individuals. In three cases we were not able to detect any mutations in any of the five exons of the TYR gene. Individuals 07 and 08 are siblings and both present the novel mutation p.C246R in heterozygous state; in addition individual 08 has the variant R402Q in heterozygous state and has a more severe albin phenotype than individual 07. We present the first report of a phenotype-genotype correlation in a Colombian OCA cohort.

3149W  
**Support to the International Rare Diseases Research Consortium: a new service to the research community.** P. Lasko1,2,2, R. Cagniard4, S. Höhn5, R. Favresse6, D. Desir-Parseille7, S. Peixoto8, M. Bellanger4, N. Lévy4, S. Aymé4. 1) International Rare Diseases Research Consortium, Montreal (IRD/IRC), Quebec, Canada; 2) CIHR Institute of Genetics, Montreal, Quebec, Canada; 3) Department of Biology, McGill University, Montreal, Quebec, Canada; 4) SUPPORT IRDiRC, Paris, France.  
The International Rare Diseases Research Consortium (IRD/IRC) brings together members that share common goals and principles and have agreed to work in a coordinated and collaborative manner within a multinational consortium to advance research in this critical area. The IRDiRC scientific secretariat has been established at the Rare Diseases Platform in Paris. IRDiRC’s two main objectives are to deliver 200 new therapies for rare diseases and means to diagnose most rare diseases by the year 2020. Its members are over 30 public and private research funding organizations that have each dedicated over 10 million US$ to research into rare diseases. A number of challenges will be addressed through collaborative actions: establishing and providing access to harmonised data and samples, performing the molecular and clinical characterisation of rare diseases, boosting translational, preclinical and clinical research, and streamlining ethical and regulatory procedures. The consortium has established three Scientific Committees. The Diagnostics committee advises on research related to the diagnoses of rare disease, including sequencing and characterization of these diseases. The Interdisciplinary committee provides expertise on cross-cutting aspects of rare diseases research including issues related to ontologies, natural history, biobanking, and registries. The therapies committee gives guidance for the pre-clinical and clinical research aiming to deliver new therapies for rare diseases. The guiding principles of IRDiRC, its plan for action and its achievements so far will be presented, as well as ways for the genetic research community to get engaged in this global effort.
3150T

Role of CFTR in regulating spermatogenesis and Implications in male infertility. H. Sharma1, R. Prasad1, S.K. Singh2, R. Mohan3. 1) Department of Biochemistry, PGIMER, Chandigarh, India; 2) Department of Urology, PGIMER, Chandigarh, India.

CFTR is a transmembrane conductance regulator (CFTR) that is a cAMP-activated chloride and HCO3 conducting channel, mutations of which are known to be associated with male infertility. CFTR mutations have been identified in patients with spectrum of genital phenotypes ranging from impaired spermatogenesis to CAVD and obstruction in reproductive tract other than vas deferens or epididymis, although some contradictory reports are available. Mutations of CFTR are identified in Indian patients with classical CF and infertile CBAVD male is heterogeneous and is completely different from that of Caucasians population, therefore Present study was undertaken to establish the spectrum and frequency of CFTR mutation in Indian infertile male population and to understand the possible involvement of CFTR gene in the etiology of male infertility other than CAVD. Blood samples from infertile males with, obstructive azospermia (n=25) with palpable vas deferens and oligospermia (n=100) were used for genomics DNA isolation and screening mutations in seven exons (2.3,4,7,8,10) hot spots of CFTR gene. Delta F-508, N1303K, R553X, G551D, G542X, 621+1G-T and W1282X were the other known most common mutation screened in Indian infertile males through allele specific ARMS PCR analysis. IVS8-Poly T allele polymorphism was also determined in all infertile male patients. Out of 25 cases with obstructive azospermia 17(68%) were found to have mutation in single CFTR allele where as among 100 oligospermic males only 17% males were also found to have mutation in CFTR gene. Delta F-508 (9.5%) and IVS8 (5T) allele (22%) were the most common mutations identified in Indian infertile males population. Twenty novel mutations in exon 10 of CFTR were also identified in patients with obstructive azospermia. The increased frequency of CF mutations in males with reduced sperm quantity and quantity and in males with azospermia without CBAVD suggest CFTR protein may be involved in the process of spermatogenesis or sperm maturation which is important to establish role of CFTR gene in the etiology of infertility. Moreover, exons 10 in NBD1 region of CFTR playing important role in etiology male infertility, therefore Indian infertile male presented with obstructive azospermia or spermatozoal defect and opting for assisted reproductive technology (ART) should be advised for screening of exon 10 of CFTR.

3151F

Milder course in Duchenne patients with nonsense mutations and no muscle dystrophin. M. Zatz, R.C.M. Pavanello, M. Lazar, N.C.V. Lourenço, A. Cerqueira, L. Nogueira, M. Vainzof. Human Genome and stem cell research center, Biosciences Institute, Universidade de Sao Paulo, Sao Paulo, SP, Sao Paulo, Brazil.

Duchenne muscular dystrophy (DMD) is a lethal X-linked condition caused by mutations in the dystrophin gene which results in the absence of muscle dystrophin. The course is usually severe and very similar in affected patients. Without careful management, loss of ambulation occurs between 10-12 years of age. On the other hand, in Becker muscular dystrophy (BMD) there is a wide variability in the severity of the phenotype which has been mainly associated with the site of the deletion and to the amount of muscle dystrophin. Therefore, the quantity/quality of muscle dystrophin has been strongly associated with appropriate muscle function. Animal models with null mutations and no muscle dystrophin are of great interest. The mdx mice are almost asymptomatic, while the GRMD (golden retriever muscular dystrophy) dogs are usually severely affected representing the best model for DMD. Such clinical difference might be related to the size of the dystrophic muscle. We have previously identified two exceptional GRMD dogs with a very mild phenotype and no muscle dystrophin. A DMD patient with a milder course despite the absence of muscle dystrophin was also reported (Dubowitz, 2006). Here we report two additional very rare DMD patients with nonsense mutations in the DMD gene and no muscle dystrophin. Case 1, currently age 14, is only mildly affected, with discrete calves’ hypertrophy, some difficulties for running and climbing stairs but with normal ability to walk. His youngest brother shows a severe course, being confined to a wheelchair at age 9. Both carry an out of frame duplication in exon 2. Case 2, is an isolated DMD patient who, at age 15, is able to walk without difficulties and climb stairs with the aid of the bannister. He carries an out-of-frame deletion encompassing exons 51 to 54. Dystrophin IF and WB with three antibodies against the N-terminal, rod domain and C-terminal, showed a typical DMD pattern, in both of them. Merosin was positive, and sarcoglycans were faint, as observed in most DMD dystrophin deficient patients. Therapeutic trials aiming the amelioration of muscle function have been focusing on the production of muscle dystrophin in affected DMD patients. Importantly, our observations indicate that it is possible to have a functional large muscle even without dystrophin. Finding the underlying protective mechanisms is of utmost importance and may lead to new avenues for treatment.

3152W

MEVF Gene Mutation Detection In Arabic Patients. R. Taha1, S. Ayesh2, M. Kambouris5, H. El-Shanti1, 2, 3. 1) Molecular Genetics, Shafallah Medical Genetics Center, Doha, Qatar; 2) Gene Medical Labs, Gaza, Palestinian Territory; 3) Yale University School of Medicine, Genetics, New Haven CT, USA; 4) University of Iowa, Pediatrics, Iowa City, IA, USA.

Autoinflammatory diseases are characterized by inflammation in the absence of high-titer autoantibodies or antigen-specific T cells. Familial Mediterranean fever (FMF) is the archetypal hereditary autosomal recessive periodic fever syndrome & auto-inflammatory disease characterized by recurrent self-limiting episodes of fever & painful polyserositis. It is found in families of Mediterranean ancestry, especially non-Askenazi Jews, Armenians, Turks, and Arabs. The offending MEVF gene localizes at Hsa 16p, encodes the pyrin (marenosmin) protein, it is highly polymorphic with multiple disease causing mutations and normal polymorphisms. In Arabic FMF patients the spectrum and distribution of MEVF mutations are distinctive and the portion of unidentified mutations (50%) is the highest amongst the groups affected by FMF. The MEVF genomic region in 100 Palestinian patients with clear FMF symptomatology consistent with the clinical diagnostic criteria and with only one identified pathogenic mutation was screened to identify the second pathogenic mutation as well as coding and non-coding variations, large duplications or deletions and intronic variations. Mutation analyses involved sequencing of exons and splice sites, sequencing putative regulatory regions by using Long range PCR, Multiplex Ligation-dependent Probe Amplification (MLPA) to detect large deletions or duplications & sequencing of the entire genomic region of MEVF. No second pathogenic mutation was identified in any of the samples by sequencing MEVF exons, splice sites, as well as putative regulatory regions. MLPA did not detect any intronic variations. Two novel rare intronic variants were identified (each in 1-3 patients) and were not present in >700 ethnically matched control chromosomes. The biological significance of these variations could not be determined. There are strong evidences of preferential effect of mutations on the others due to extensive polymorphism within the genomic sequence that would account for the lack of detection of the second pathogenic mutation. Alternatively, the effects of modifier genes or other loci that influence the clinical picture of FMF in Arabic population also seems to be involved. The comprehensive identification of MEVF mutations alleles among FMF patients is essential for the efficient examination of specific genotype-phenotype correlation patterns and for development of molecular tools to support the clinical diagnosis.
PhenomeCentral: An Integrated Portal for Sharing and Searching Patient Phenotype Data for Rare Genetic Disorders. M. Brudno1; M. Girdea1; S. Domítriu1; S. Kohler3; P.N. Robinson2; A.J. Brookes3; K. Boycott1; C.F. Boerkoel2; W.A. Gahl1, Canadian CARE for RARE Consortium (FORGE) and NIH Undiagnosed Diseases Program. 1) Centre for Computational Medicine, Hospital for Sick Children & University of Toronto, Toronto, ON, Canada; 2) Charité Hospital, Berlin, Germany; 3) Department of Genetics, University of Leicester, Leicester, United Kingdom; 4) Children’s Hospital of Eastern Ontario, Ottawa, ON Canada; 5) National Institutes of Health Undiagnosed Diseases Program.

The availability of low-cost genome sequencing has allowed for the identification of the molecular cause of hundreds of rare genetic disorders. Solved disorders, however, only represent the ‘tip of the iceberg’. Because the discovery of disease-causing variants typically requires confirmation of the mutation or gene in multiple unrelated individuals, an ever larger number of genetic disorders remain unsolved due to difficulty identifying second families. Many with groups now tackling these remaining undiagnosed disorders, which may be present in only a handful of individuals seen at different hospitals and sequenced by different centers, it is critical to establish effective and secure data-sharing techniques that allow clinicians and scientists to identify additional families via phenotype and genotype searches.

To address this need, we have developed PhenomeCentral (http://phenome-central.org), a repository for secure data sharing targeted to the rare disorder community. Each patient record within PhenomeCentral consists of a thorough phenotypic description capturing observed abnormalities as well as relevant absent manifestations, expressed using Human Phenotype Ontology terms. Furthermore, each record can be labeled by the creator private, hidden from everyone except the contributor; public, viewable and searchable by all registered users; or matchable the record cannot be directly viewed or searched, but is reachable via an automated phenotype matching system (following Cafe Variome principles) which informs contributors of the existence of profiles similar to their cases. The phenotypic features shared among these records are presented without revealing additional patient information or the contributors, enabling direct communication for any subsequent data sharing.

PhenomeCentral currently incorporates phenotype data for >400 patients with rare genetic disorders without a molecular diagnosis, including: 200 from the Canadian CARE for RARE project and 150 from the NIH Undiagnosed Diseases Program (UDP). Clinical geneticists and scientists studying rare disorders can request access accounts, and new patients can be added further using the PhenoTips User Interface, built into PhenomeCentral, or uploaded in bulk. An interface allowing for the deposition and analysis of genomic data (whole-exome VCF files and CNVs) is under development.

Ocular phenotypes in aneurysm syndromes collected from GenTAC (National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions) Registry. G. Oswald1,2, R. Parvari1,4, S. Köhler2, P.N. Robinson2, A.J. Brookes3, K. Boycott1, E. Reynolds3, R. Yerushalmi1, M. Arafat1,4, M. Brudno1, R. Ravekes1, W. A. Gahl1, B. Yerushalmi1, M. Oswald1, S. Milo1, EM. Reynolds2, R. E. Pyeritz1, R. Devereux1, DM. Miliewicz1, E. M. Reynolds2, JP. Habashi2, GenTAC Registry Consortium. 1) Johns Hopkins University, Baltimore, MD; 2) Baylor College of Medicine, Houston, TX; 3) NIA at Harbor Hospital, Baltimore, MD; 4) Oregon Health & Science University, Portland OR; 5) Queen’s Medical Center, Honolulu, HI; 6) The University of Pennsylvania, Philadelphia, PA; 7) Weill Cornell Medical College of Cornell University, New York, NY; 8) University of Texas Medical School at Houston, Houston, TX; 9) University of Maryland, Baltimore, MD.

The National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions (GenTAC) has enrolled over 3400 patients with aortic aneurysm syndromes, including Marfan (MFS), Loeys-Dietz (LDS), vascular Ehlers-Danlos (vEDS), Bicuspid Aortic Valve with enlargement (BAVe) and Familial Thoracic Aortic Aneurysm Disease (FTAAD). The GenTAC Registry was queried for prevalence of ocular features, showing its utility to investigators studying non-cardiac features of these disorders. Ectopia lentis (ECL) was reported in 33% of patients with MFS, with equal distribution between age and gender, consistent with previously published data. ECL was not observed in LDS or vEDS, and rarely reported in other aneurysm conditions (+1%). Retinal detachment (RD) was observed in 8.7% of MFS and rarely in other aneurysm syndromes (+1%). ECL and RD were concomitant in 51 of 61 MFS patients. Glaucoma and cataracts were reported in 4% and 8.3% of MFS patients, respectively. Myopia was most commonly reported in MFS, but not significantly different than that observed in the general population. LDS showed the highest prevalence of strabismus/amblyopia of 12.3% as compared to MFS (8.3%) and other syndromes. Patients with MFS are commonly counseled to avoid LA/SIK correction of refractive error due to speculation that surgery’s mechanical forces may increase RD risk. Analysis of the operative data available in a subset of patients found that 3.1% of enrolled patients with a confirmed diagnosis of MFS reported having undergone LA/SIK. Of the 12 MFS patients who had LA/SIK, 7 had ECL (58.3%) and 2 had RD (16.7%). 2/12 patients who had had LA/SIK previously reported therefore, presence or absence of lens dislocation is a useful diagnostic tool in differentiating patients. Furthermore, this data supports the theoretical risk that LA/SIK correction may increase the risk of retinal detachment in patients with connective tissue disorders however more detailed studies investigating the temporal relationship need to be done.
In silico and molecular analyses of mutations that alter mRNA splicing of COL1A1. J. Schlett, S. Bailey, T. Tran, D. Chen, P.H. Byers. Pathology, University of Washington, Seattle, WA.

Approximately 10-20% of inherited pathogenic mutations alter mRNA splicing. Phenotypes of affected individuals are often dependent on the stability of the mRNA produced by the mutant allele. Currently, the downstream effects of mutations on RNA splicing are often difficult to predict, complicating diagnoses based on DNA sequence alone. As the use of molecular diagnostics moves increasingly to DNA sequencing, the need to understand the effects of splice site mutations will increase further. We sought to identify factors that determine the mRNA splice products in the heritable disorder osteogenesis imperfecta (OI). To better understand the relationship between genotype and mRNA splicing outcomes, we examined clinical and molecular data from 219 OI patients with 122 unique mutations within the introns COL1A1, which encodes the proa1(1) chain of type I collagen. We examined mutations representing OI types I-IV with approximately 73% of patients diagnosed with OI Type I, 11% with OI Type IV, 6% with OI Type III, and 10% with OI Type II. Of the 219 (48%) of patients had mutations in the canonical GT-AG splice donor or acceptor site. All patients with OI Type I, for which cDNA was examined, produced unstable mRNAs with premature termination codons due to translational frameshifts. 34 individuals with G to A transitions in the final nucleotide of COL1A1 introns (IVS(X)+1G>A) had OI type I phenotypes, due to shift of the splice acceptor site one nucleotide downstream when the first nucleotide of the exon was guanine. In contrast, 72 individuals with IVS(X)+1G-A mutations had highly variable OI phenotypes (Types I-II) and the splice outcomes yielded diverse and often multiple mRNA products. In patients with OI Types I-II we detected abnormal, apparently mis-spliced mRNAs. For splice donor site mutations there was no correlation between outcome and intron size or the predicted strength of native and alternative splice sites in these individuals. To examine additional factors which may impact splice outcome, we treated wild-type human fibroblasts with the transcriptional inhibitor Actinomycin D. Using intron specific primers, we measured the speed and order of intron splicing of COL1A1 using quantitative PCR and capillary electrophoresis. mRNA splice outcomes were independent of the speed at which mRNAs were synthesized but the order in which introns are normally removed appears to play a role in determining splice outcome .


Chitinase hydrolyze chitin, a polymer of N-acetyl-D-glucosamine, which is normally removed appears to play a role in determining splice outcome .

Genotype and Family Analysis of 68 Thai families with Duchenne Muscular Dystrophy. L. Choubum1, K. Taweechue2, W. Khunin3, S. Nujarean3, C. Limwongse3, D. Wattanasirichai3. 1) Research Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 2) Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; 3) Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 4) Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.

Background: Duchenne muscular dystrophy (DMD; OMIM#300377) is a severe X-linked devastating disorder, resulted from mutation in the dystrophin gene (DMD) located on Xp21.2-p21.1. It contains 79 coding exons and encodes a 3,685 amino acid residues-containing peptide. The aim of this study is to describe genotypic defects and family structure including positive family history, carrier rate, and population screening. Methods: Patients with the diagnosis of DMD based on clinical manifestation, family history and elevated levels of creatine kinase were enrolled in the study through genetics/neuromuscular clinics and network for prevention of DMD (the Thai Project, the Thai Project, Faculty of Medicine Ramathibodi Hospital. Multiplex PCR and/or multiple ligation-dependent probe amplification (MLPA) were employed as the first tier genetic analysis of the DMD gene. If the result of the MLPA showed abnormal, then it was followed by PCR-direct sequencing for all 79 coding exons and brain/muscle promoter region of the gene. Results and Conclusion: A total of 68 unrelated DMD families were enrolled in the study. Multiplex PCR/MLPA revealed 72% of the probands harboring exon(s) deletion (41) or duplication (8). Nineteen probands had sequence analysis, 10 of whom the result are available for families, specific including nonsense and frameshift mutations. There were positive family history in 17 (22%) families, and family data was incomplete in 8 families. There were a total of 115 affected (deceased and living) cases with the maximum affected case of 8 in one family. Thirty-three mothers of the 43 probands with negative family history were tested for carrier status, and 39% (12 out of 31) were shown to be carrier. The results of this study support the previous observation that large deletion/duplication of the DMD gene may account for two thirds of cases of the DMD. In this study, there was a lower rate of positive family history and maternal carrier for sporadic cases compared to those previously described in the literature.
Novel mutations in EDA gene in hypodontia and curly hair. J.-W. Kim¹,², K.-E. Lee¹, J. Ko¹. ¹) Pediatric Dentistry, Seoul National University School of Dentistry, Seoul, South Korea; ²) Molecular Genetics, Seoul National University School of Dentistry, Seoul, South Korea.

Hypodontia (tooth agenesis) is the developmental absence of at least one tooth except third molar. Familial hypodontia can occur as an isolated form or as part of a genetic syndrome. Mutations in MSX1, PAX9 and AXIN2 genes has been identified in familial non-syndromic hypodontia. Ectodermal dysplasia is a group of syndromes affecting ectodermal origin tissues and comprises more than 150 different forms. Mutations in the ectodysplasin-A (EDA) gene have been associated with X-linked hypohidrotic ectodermal dysplasia and partial disruption of the EDA signaling pathway has been shown to cause isolated form of hypodontia. We have identified two X-linked hypodontia families, and performed mutational analysis of EDA gene. Mutational analysis revealed two novel EDA mutations: c.866G>T, p.Arg289Leu; c.1135T>G, p.Phe379Val (reference sequence NM_001399.4). These mutations were perfectly segregated with hypodontia and curly hair within each family and were not found in 150 control X-chromosomes with same ethnic background and 1000 genome project. This study broadens the mutational spectrum of EDA gene and understanding of X-linked recessive hypodontia with curly hair. This work was supported by grant (02-2013-0002) from the SNUDH Research Fund and by grants from the Bio & Medical Technology Development Program (2011-0027790), the Science Research Center grant to Bone Metabolism Research Center (2012-0000487) by the Korea Research Foundation Grant.
3162W

Mutation in the mouse homolog of C5ORF42 disrupts cilogenesis and causes cerebellar defects and other Joubert Syndrome phenotypes associated with the disruption of SHH signaling. R. Damerla1, C. Oul2, G. Gabrielli1, X. Liu1, B. Gibb1, R. Francis1, Y. Li2, B. Chatterjee1, M. Srou2, J.L. Michaud1, G. Pazour2, C.W. Lo1. 1) Developmental Biology, University of Pittsburgh, Pittsburgh, PA; 2) Centre of Excellence in Neurosciences of Université de Montréal and Sainte-Justine Hospital Research Center, Montreal, Quebec, Canada H3T 1C5; 3) Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

Mutations in cilia-related genes are observed to cause a host of rapidly expanding genetic disorders known as ciliopathies. Two recent studies of patients in Canada and Saudi Arabia linked mutations in an uncharacterized gene C5ORF42 with Joubert syndrome (JSRD). We report for the first time a mouse model for the Joubert syndrome with a missense mutation in 2410089E09Rik, the mouse homolog of C5ORF42. This mutant, named Heart Under Glass (Hug) was recovered from a forward genetic screen with ENU mutagenesis using fetal ultrasonic imaging to identify congenital heart defects. Genome scanning to map the mutation followed by exome sequencing analysis identified the mutation in Hug as c.T757C:p.S235P. Hug mutants exhibited a host of defects characteristic of JSRD patients, including polydactyly, craniofacial anomalies, cystic kidneys, and congenital heart defects comprising of outflow tract malalignment associated with pulmonary atresia. These developmental anomalies are also reminiscent of other ciliopathies, suggesting they may arise from a ciliogenesis defect and the disruption of cilia-dependent planar cell polarity pathways (PCP) and Shh signaling. Immunocytochemistry in fibroblasts derived from both the ENU mutant mice and from a JBS patient with a C5ORF42 mutation showed defects in cilogenesis. This was associated with perturbation of the cilia transition zone. Shh signaling was observed to be disrupted in Hug mutant fibroblasts, but stereocilia patterning was not perturbed, suggesting cochlear PCP regulation is not affected. Detailed analysis of the brain showed abnormal cerebellar development associated with a significant reduction in vermis foliation, defects that are clinically relevant for assessing JBS prognosis. Together, these studies showed the novel Hug mutant mouse model will be invaluable for further interrogating the role of cilia biology in the pathophysiology of JBTs.
3165W
Investigating the clinical features and genetics of idiopathic generalized epilepsy starting in mothers of babies with epilepsy. S. Ghavimi1, H. Azimi2,3. 1) Mofid Hospital, Shahid Beheshti Medical University, Tehran, Iran; 2) All Saints University School of Medicine, Dominica; 3) PsychoGenome, Ottawa, Ontario, Canada.

Objective: To investigate the clinical features and genetics of idiopathic generalized epilepsy starting in mothers of babies with epilepsy. Methods: Patients with general spike, defined as generalized seizures with spike or polyspike and wave on EEG, were studied in the setting of a first seizure clinic where an early postictal EEG record is part of the protocol. This outpatient study was conducted on 200 mothers who came to the clinic of Mofid's Hospital. Results: Of 200 mothers with an electro-clinical diagnosis of IGE, 56 (28%) were diagnosed as adult onset IGE. The seizure patterns in these 30 cases were tonic-clonic seizures + absences (8), tonic-clonic seizures alone (45). Tonic-clonic seizures were often precipitated by alcohol or sleep deprivation. The proportion of affected first and second degree relatives did not differ between the classical and adult onset IGE groups. Twenty adult onset cases were treated with sodium valproate, four with other antiepileptic drugs and two were untreated. Conclusions: Babies born to mothers who are epileptic can give an increase in chance of being epileptic themselves. Although it is still not clear how or when the epilepsy will start, it is clear that these children, just like their mothers carry the genes, and a mutation required to have idiopathic generalized epilepsy. Adult onset IGE is a relatively frequent and benign disorder. Seizures are usually provoked and are easy to control. Patients in this age group may often be misdiagnosed as having non-lesional partial epilepsy. Early postictal EEG and sleep deprivation studies may improve the detection of these patients. Pedigree analysis suggests that adult onset IGE, like classical IGE, has a genetic etiology.

3166W
The role of fibrillin-1 in human mesenchyme stem cell adipogenesis. M.R. Davis1, C. Duffy2, P. DeSousa2, V. MacRae1, K.M Summers1. 1) Genetics and Genomics, The Roslin Institute, The University of Edinburgh, Roslin, Midlothian, United Kingdom; EH25 9RG; 2) MRC Centre for Regenerative Medicine, SCRM Building, The University of Edinburgh, Edinburgh BioQuarter, 5 Little France Drive, Edinburgh, EH16 4UJ.

The extracellular matrix is important in maintaining the structure of connective tissues, including bone, skin, blood vessels and adipose. Fibrillin-1 is a leading component of 10 nm microfibrils, which provide strength and elasticity to connective tissues. Mutations in the fibrillin-1 gene lead to the connective tissue disorder Marfan syndrome (MFS), which affects the cardiovascular and musculoskeletal systems. Many patients suffering from MFS present with depletion in adipose tissue throughout their bodies, which has also been demonstrated in mouse models. Therefore it is important to further investigate the significance of fibrillin-associated microfibrils in adipose formation. This study used mesenchymal stem cells (MSCs), which are capable of differentiating into multiple connective tissue lineages, to investigate the role of fibrillin-1 in formation of adipocytes. Fibrillin expression at the mRNA and protein level was examined using fluorescent immunocytochemistry, quantitative PCR (qPCR) and bioinformatics validation utilizing various online databases. The study demonstrated the presence of fibrillin-1 microfibrils and RNA expression early in primary human MSC differentiation to the adipose lineage. Fibrillin-1 was degraded and mRNA levels decreased as differentiation proceeded. Since many MFS patients lack appropriate formation of adipose tissue, we suggest that a fibrillin matrix is necessary for the early stage of differentiation into this connective tissue lineage but not required to develop the specific differentiated state.
Heritability of obesogenic growth trajectories during development in a model system. C. Schmitt1, S. Service2, R. Cantor3, A. Jasinska4, M. Jorgensen3, J. Kaplan5, N. Freimer6, 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA; 2) Department of Human Genetics, UCLA, Los Angeles, CA; 3) Department of Pathology, Section on Comparative Medicine, Wake Forest University School of Medicine, Winston-Salem, NC.

Obesity is increasingly prevalent worldwide, and has severe negative effects in public health. Obesity arises from a complex interaction of genetic predisposition and environment that can accumulate throughout life. Although obesity risk is known to increase with age, the markers contributing to obesity in the manifestations of adult disease, few studies have been undertaken on the developmental measurements that might be associated with adult obesity risk. The search for obesgenic markers during development in humans is complicated by the ubiquity of diets high in fat and simple carbohydrates, and the difficulty in assessing the actual diets of study subjects. This research investigates the genetic underpinnings of obesogenic growth trajectories from birth to adulthood in a genetically well-characterized model system under a controlled diet and environment: the African green monkey (Chlorocebus aethiops sabaeus) in the Vervet Research Colony at Wake Forest School of Medicine.

We used growth curve analysis on measures taken thrice yearly on body size and composition - body weight (BW), BMI and waist circumference (WC) - in a population of 641 monkeys measured from 2000 through 2012, 33 individuals, 6 M and 27 F, presented with signs of chronic abdominal obesity - defined as having an adult WC above 40.5 cm for at least three successive measurements. Individuals measured < 6 times were excluded from analysis. Rate of growth (h = 0.54, p < 0.0001) of the BW growth curves fitted by threeparameter logistic growth curves in nonlinear mixed effects models, with parameters modeled as fixed effects and subject within sex and obesity status modeled as random effects. As expected, we found a strong effect of sex and a trend toward obesity status on all parameters of growth. We assessed heritability of individual growth parameters using maximum likelihood variance components analysis in SOLAR. Growth parameters were highly and significantly heritable, with sex as a significant covariate (e.g., BW, h² = 0.32, rate of growth, h² = 0.34, p < 0.0001; WC, h² = 0.30, rate of growth, h² = 0.34, p < 0.0001). This study suggests that adult obesity is a developmental process driven in part by heritable obesogenic trajectories resulting in faster and longer growth to larger adult size. A more detailed examination of individual ancestry will be used to better understand the heritable obesity risks and promote the discovery of novel biomedical interventions.

Heritability of obesogenic growth trajectories during development in a model system. C. Schmitt1, S. Service2, R. Cantor3, A. Jasinska4, M. Jorgensen3, J. Kaplan5, N. Freimer6, 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA; 2) Department of Human Genetics, UCLA, Los Angeles, CA; 3) Department of Pathology, Section on Comparative Medicine, Wake Forest University School of Medicine, Winston-Salem, NC.

Obesity is increasingly prevalent worldwide, and has severe negative effects in public health. Obesity arises from a complex interaction of genetic predisposition and environment that can accumulate throughout life. Although obesity risk is known to increase with age, the markers contributing to obesity in the manifestations of adult disease, few studies have been undertaken on the developmental measurements that might be associated with adult obesity risk. The search for obesgenic markers during development in humans is complicated by the ubiquity of diets high in fat and simple carbohydrates, and the difficulty in assessing the actual diets of study subjects. This research investigates the genetic underpinnings of obesogenic growth trajectories from birth to adulthood in a genetically well-characterized model system under a controlled diet and environment: the African green monkey (Chlorocebus aethiops sabaeus) in the Vervet Research Colony at Wake Forest School of Medicine.

We used growth curve analysis on measures taken thrice yearly on body size and composition - body weight (BW), BMI and waist circumference (WC) - in a population of 641 monkeys measured from 2000 through 2012, 33 individuals, 6 M and 27 F, presented with signs of chronic abdominal obesity - defined as having an adult WC above 40.5 cm for at least three successive measurements. Individuals measured < 6 times were excluded from analysis. Rate of growth (h = 0.54, p < 0.0001) of the BW growth curves fitted by threeparameter logistic growth curves in nonlinear mixed effects models, with parameters modeled as fixed effects and subject within sex and obesity status modeled as random effects. As expected, we found a strong effect of sex and a trend toward obesity status on all parameters of growth. We assessed heritability of individual growth parameters using maximum likelihood variance components analysis in SOLAR. Growth parameters were highly and significantly heritable, with sex as a significant covariate (e.g., BW, h² = 0.32, rate of growth, h² = 0.34, p < 0.0001; WC, h² = 0.30, rate of growth, h² = 0.34, p < 0.0001). This study suggests that adult obesity is a developmental process driven in part by heritable obesogenic trajectories resulting in faster and longer growth to larger adult size. A more detailed examination of individual ancestry will be used to better understand the heritable obesity risks and promote the discovery of novel biomedical interventions.
3172W

Mouse models reveal an essential role for RERE in eye development. B. Kim1, O. Shchelochkov1, M. Justice1, B. Lee1, D. Scott1. 1) Molecular and Human Genetics, Baylor College Med, Houston, TX; 2) Department of Pediatrics, University of Iowa, Iowa IA.

Microphthalmia occurs in approximately 1 out of 10,000 individuals and can be caused by alterations in genes involved in the early development of the eye. In an effort to identify novel genes involved in the development of microphthalmia, we carried out an autosomal recessive ENU screen. We found a novel mouse strain (eyes3) with microphthalmia. The mutation responsible for the eyes3 phenotype was mapped by linkage analysis to a region of mouse chromosome 4 that is syntenic to human chromosome 1p36.31-p36. Rere (arginine-glutamic acid dipeptide repeats) maps to this region and was selected as a positional candidate based on its role as a nuclear receptor co-regulator. Sequencing revealed a homozygous c.578T>C change in Rere, which produces a single amino acid change in a highly conserved BAH domain of Rere (p.Val193Ala). To confirm that the microphthalmia seen in the eyes3 strain was due to a defect in Rere, these mice were crossed with mice carrying an Rere null allele (om). Homozygous om embryos (Rere<sup>om/om</sup>) die around E9.5, but a portion Rere<sup>om/om</sup> mice lived into adulthood but have microphthalmia and optic nerve atrophy. We examined the expression pattern of Rere in the developing mouse eye. At E13.5, RERE expressing cells were primarily located in the lens epithelial cells and the optic cup margin. RERE was detected in the ganglion cell layer and the lens at birth. At postnatal day 14 (P14), RERE was expressed in the outer nuclear layer (ONL), inner nuclear layer (INL), and ganglion cell layer (GCL). Histological examination revealed that the lens was not formed in Rere<sup>om/om</sup> embryos at E15.5. In wild type embryos, invagination of the lens placode results in formation of the optic cup and the lens vesicle around E10.5. In Rere<sup>om/om</sup> embryos, the lens vesicle is not developed due to failure of invagination of the lens placode which, consequently, leads to shallow lens pit and abnormal development of the optic cup. This is consistent with ERGs showing retinal rod and cone signaling which is required for reciprocal interactions between the optic vesicle and invaginating lens placode. We conclude that RERE is required for lens induction in mice. It is likely that RERE performs a similar role in humans and may contribute to the mouse phenotypes through its effects on retinoic acid signaling in the developing eye.

3173W

Functional and mutational analysis of long-range enhancers of ZIC3 in patients with congenital heart defect and laterality in zebrafish models. J. Marino1,2, S. Hook1, P. Hu2, R. Hart2, E. Rossler2, J.A. Towbin2, J.W. Belmont3, L. Ribeiro-Bicudo4, M. Muenke2. 1) Genetics, Hospital for Rehabilitation of Craniofacial Anomaly Rua Silvio Marchioni, 3-20 Vila Universitária 17012-900 - Bauru, SP - Brasil - Caixa-postal: 1501; 2) National Institute of Health, Bethesda, MD, USA National Institutes of Health 35 Convent Drive, Bldg 35, Room 1B202 Bethesda, MD 20892-3717; 3) Baylor College of Medicine, Houston, TX, USA.

Mutations in ZIC3 frequently result in X-linked heterotaxy in humans, a syndrome consisting of left-right patterning defects, midline abnormalities, and cardiac malformations. Studies in mouse models of Zic3 dysfunction also result in heterotaxy, indicating conserved mammalian function of this developmental transcription factor. Deletion of the entire ZIC3 locus in humans, or gene dysfunction in the classical mouse mutant bent tail model, results in heterotaxy, indicating that loss-of-function is the most common pathogenic mechanism. Here we use comparative genomic alignments of non-coding elements near ZIC3 to identify and test for conserved function of potential human regulatory elements in the zebrafish model. We set out to examine the ZIC3 genomic locus by testing all of the potential enhancers and promoter elements identified by the ECR browser tool (http://ecrbase.dcode.org). Functional elements were then subjected to Sanger sequencing in 380 control samples compared with 366 patients with congenital heart defects and/or laterality. We now describe two mutations in the ECR1 within a 521bp fragment, a C-G transition and G-C transition (both in affected males). Similarly, ECR2 showed a male specific C-A transition mutation in a 433bp functional enhancer fragment. These findings are in addition to several other mutation positive regulatory elements at the ZIC3 locus. Here we describe both the functional and mutational landscape of the laterality gene ZIC3 and provide evidence that non-coding elements will likely prove to be important in our understanding of disease pathogenesis and in future genetic testing.

3174W

Functional characterization of Gli2 in normal breast development and in breast cancer. C. Zhao, PA. Beachy. Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, CA, 94305.

The highly conserved Hedgehog (Hh) signaling pathway plays critical roles in embryogenesis and adult tissue homeostasis. Gli2, an essential transcriptional factor of the Hh pathway, is required for embryonic development and postnatal growth of various tissues. Germline mutations in Gli2 often lead to developmental disorders, such as holoprosencephaly or hypopituitarism; whereas somatic mutations or gene amplifications of Gli2 have been found in a growing list of malignancies, including breast cancer. Despite progress in identifying germline or somatic Gli2 mutations in ectoderm-derived tissues, such as in the skin, little is known about function of Gli2 in mammmary gland development and in mammmary tumors. We find that Gli2 is expressed in mammmary stromal cells and its expression is differentially regulated via secreted factors from mammmary epithelial cells during a mammmary regenerative cycle. Importantly, generic ablation of Gli2 function in mammmary stromal cells results in a distended mammmary gland with disrupted terminal end buds (TEB). Intriguingly, a subset of Gli2 mutations in human breast cancer when expressed in FACS-isolated mammmary stromal cells causes elevated Hh pathway activity. We propose a paracrine model that involves epithelial-secreted factors to regulate stromal-specific Gli2 expression in normal mammmary development and deregulation of this epithelial-stromal interaction contributes to mammmary malignancies.

3175W

Modeling Foxf1 deficiency and overexpression in mice. A.V. Dharmadhikari1,2, B. Caroño1,2, M.G. Hill1, X. Ren1, T.V. Kallin2, J. Zabielska4, W.Y. Wari2, T. Majewski5, H.B. Brown1, A. Gambin3, P. Szafrański3, V.V. Kalinichenko1, M.J. Justice1,2, P. Stankiewicz1,2. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Program in Translational Biology & Molecular Medicine, Baylor College of Medicine, Houston, TX; 3) Division of Pulmonary Biology, Cincinnati Children’s Hospital Research Foundation, Cincinnati, OH; 4) Mossakowski Medical Research Centre, Polish Academy of Sciences, Warsaw, Poland; 5) Dept of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX; 6) Dept of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX.

Alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV) is a rare neonatally-lethal diffuse developmental disorder of the lungs caused by haploinsufficiency of Foxf1. All affected newborns die in the first month of life due to severe respiratory distress and pulmonary hypertension. Foxf1 null mice die by midgestation as a result of defects in mesodermal differentiation and cell adhesion. Foxf1 heterozygous mice exhibit up to 90% neonatal mortality, depending on genetic background. For the current study, Foxf1 heterozygous mice with a deletion of the forkhead binding domain were generated and are congenic on the C57BL/6J background. In contrast to recent reports that FOXF1 is incompletely paternally imprinted in the human lungs, early postnatal mortality was observed regardless of parental transmission of the deleted allele; no differences were seen in Foxf1 expression in embryonic and postnatal lung tissues from reciprocal crosses. Analysis of RNA from postnatal day 0.5 Foxf1<sup>−/−</sup> and wildtype lungs using Illumina mouse WG-6 v2.0 expression bead chip microarray revealed statistically significant deregulation (p<0.05, fdr<0.05) of several genes, including those involved in pulmonary vascular development (Sema3C, Dil4, and Ednrb), lung branching morphogenesis (Fgf10 and Lama1), and the blood pressure regulating renin-angiotensin system (Ren1, Cma1, and Cpa3). To study the effects of Foxf1 overexpression, we knocked a Cre-inducible Foxf1 allele into the ROSA26 locus. These mice have been mated to CMV-cre and Tie2-cre mice to obtain whole body and vascular endothelial cell specific overexpression of Foxf1, respectively. We are currently mating ROSA26Foxf1; Tie2-cre mice to Foxf1<sup>−/−</sup> mice, which we hypothesize will rescue early postnatal mortality. This could inform future gene therapy studies in patients with ACDMPV.
3176W
Transgenic zebrafish expressing mutant skeletal muscle actin acta1a genes model human nemaline myopathy. O. Ceyhan, A.H. Beggs. Division of Genetics and Program in Genomics, The Mount Center for Orphan Disease Research, Boston Children's Hospital and Harvard Medical School, Boston MA, USA.

The nemaline myopathies (NMs) are a group of rare genetic neuromuscular disorders that tend to present at birth or infancy with moderate to severe muscle weakness and defined by an accumulation of rod-like structures (nemaline bodies) in myofibers. Heterozygous (dominant) mutations in the skeletal muscle α-actin gene (ACTA1) account for ~25% of all NM cases and ~50% of the severe presentations. Despite our current understanding of normal actin function, the mechanisms that lead to defects in muscle development and function in patients with ACTA1 mutations remain unclear and there are no curative therapies available for treatment. In order to elucidate the molecular defects underlying the muscle pathology in ACTA1-related NM, we are generating a panel of transgenic zebrafish lines that express a series of disease-linked dominant ACTA1 mutations on the zebrafish acta1a transcript, the predominant actin expressed in the majority of zebrafish skeletal muscle. We initially tested whether three of these variants, acta1a H42Y, M134V, and V165M, would lead to a muscle phenotype by injecting the mutant mRNAs into zebrafish embryos. Fish overexpressing mutant actins displayed myopathic phenotypes characterized by delayed hatching from the chorion and curved bodies to variable degrees that correspond with the severity of disease in patients with these mutations. We next generated an acta1a V165M transgenic line that stably expresses mutant actin in all fast myofibers, and characterized its neuromuscular phenotypes by morphological and histological analysis at different time points. The acta1a V165M fish display muscle weakness as evidenced by their thin bodies and curved tails at 2 days-post-fertilization (dpf) and reduced motility in touch-evoked escape response assay at 5 dpf. Whole-mount phalloidin staining at 2 dpf revealed actin aggregates in the affected fish muscle, while electron microscopy demonstrated Z-line thickening and severe myofibrillar disorganization. These results provide proof of concept that zebrafish models of actin mutations recapitulate the human disease and have robust myopathic phenotypes that may be amenable for high-throughput chemical screening.

3177W
Zebrafish ptk7 loss-of-function mutants reveal useful genetic models for human congenital and idiopathic scoliosis. M. Hayes1,2, B. Ciruna1,2. 1) Program in Developmental & Stem Cell Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Ontario, Canada.

Scoliosis refers to three-dimensional curvatures of the spine and is typically broken-down into two sub-categories: congenital scoliosis (CS) that presents with vertebral anomalies associated with developmental defects, and idiopathic scoliosis (IS) diagnosed by curvature with no underlying abnormality. IS typically develops postnatally, most commonly in adolescents. Despite decades of research into the causes of scoliosis, the lack of appropriate animal models has made the investigation of possible pathological mechanisms difficult. We have generated loss-of-function ptk7 mutant zebrafish and have observed a highly penetrant spinal curvature phenotype. Maternal-zygotic (MZ) ptk7 mutant embryos, which represent a complete loss-of-function model, display non-canonical Wnt/β-catenin signaling defects. Ptk7 expression in the tail bud and defects in Wnt/β-catenin-dependent mesodermal specification in MZptk7 suggest a role for ptk7 in vertebral patterning in vivo. Indeed, we observe vertebral malformations in MZptk7 larvae that closely mimic human CS phenotypes. We suggest a role for early ptk7 expression and Wnt/β-catenin signaling in axial development and CS pathology. Interestingly, zygotic loss of ptk7 does not affect vertebral patterning and we do not observe vertebral malformations in mutant larvae. However, ptk7 mutant juveniles develop severe axial curvatures that progress up until sexual maturity. We suggest that with no underlying vertebral abnormalities, ptk7 mutant zebrafish represent a model of adolescent idiopathic scoliosis (AIS) and we are currently using these zebrafish to investigate the mechanisms involved in pathogenesis. Our evidence suggests that depending on the timing of loss-of-function, similar gene pathways may be associated with both CS and IS. Common pathological mechanisms may be associated with both types of abnormalities and we are using loss of ptk7 as a model to test possible contributing factors to the human disease.

3178W
FGF Ligands Regulate Chondrocyte Differentiation in the Proximal Limb. I.H. Hung1,2, D.M. Ornitz2, G.C. Schoenwolf2, M. Lewandoski2. 1) Pediatrics/Medical Genetics, University of Utah, Salt Lake City, UT; 2) Cancer and Developmental Biology Lab, National Cancer Institute, Frederick, MD; 3) Dept. of Developmental Biology, Washington University School of Medicine, St. Louis, MO; 4) Dept. of Neurobiology and Anatomy, University of Utah, Salt Lake City, UT.

Activating mutations in fibroblast growth factor (FGF) receptors result in chondrodysplasia and craniosynostosis syndromes, highlighting the critical role for FGF signaling in skeletal development. Although the roles of the FGFs in bone development have been relatively well-characterized, only two FGF ligands, FGF9 and FGF18, have been shown to regulate embryonic skeletogenesis. Our previous analyses of FGF9 and FGF18 single knock-out mice suggested that these FGF ligands may be functionally redundant. To test this hypothesis and to further elucidate their roles, we have generated and analyzed the limb phenotypes of an Fgf allelic series. Here we demonstrate novel roles of these FGF ligands in chondrogenesis of the proximal limb.

3179W
The maternal polymorphism rs2236131 in ITPK1 gene is associated with neural tube defects in a high-risk Chinese population. z. Guan, J.H. Wang, J. Guo, F. Wang, XW. Wang, GN. Li, Q. Xie, X. Han, B. Niu, T. Zhang, Wang JH. Department of biotechnology, Capital Institute of Pediatrics, Beijing, Beijing, China.

Neural tube defects (NTDs) are common and severe malformations that are multifactorial, involving the combined action of both genetic and environmental factors. Insitol as maternal nutritional facts and its related genes are suggested to be implicated in Neural tube defects (NTDs), but the mechanisms are not clear. Insitol, 1,3,4-trisphosphate 5/6-kinase (ITPK1) is a key enzyme in inositol metabolism, and has been studied for gene mutation in the mouse but not single nucleotide polymorphism (SNP) in NTD-affected pregnancies. A case-control study of women with NTD-affected pregnancies (n=200) and controls (n=320) from a high-risk area for NTDs in China was carried out to investigate the association of the polymorphisms in ITPK1 gene with NTDs. The 13 tag SNPs of ITPK1 chosen based on the minor allele frequency (MAF) ~20% were genotyped by the Sequenom MassArray system, and we found that 4 tag SNPs were statistically associated with NTDs (p<0.05). After stratifying participants by NTD phenotypes, the significant association only existed in cases with spina bifida. We predicted the binding capacity of transcription factors in the 4 tag SNPs using the bioinformatics method. Only the rs2236131 is located in the sequence transcription factors (SP-1). EMSA (electrophoretic mobility shift assay) was applied to verify the binding activity between wild and mutated oligonucleotides probes for the positive SNP (rs2236131), and showed a different allelic binding capacity of specificity protein-1 in the intron region of the ITPK1, which is affected by an G→A exchange. The RT-PCR showed that the expression decreased significantly in mutant type with rs2236131 compared with wild type in the health pregnancy (P<0.05). These results suggested that the maternal polymorphism rs2236131 of ITPK1 was a potential risk factor for NTDs in a high-risk area of China and the allele A of rs2236131 in ITPK1 might affect the ITPK1 gene expression level. These can supply one of the worthwhile predictor of NTDs. Foundation: National Natural Science Foundation (81070491).
SPECC1L deficiency causes neural crest cell delamination and migration defects in orofacial clefting. I. Saadh, N. R. Wilson, A. J. Ohm-Shipman, E. Kosa, D. S. Acuña-Rueda, K. M. Stumpff, G. Smith, L. Pittick, B. C. Bjork, A. Czirok. 1) Dept. of Anatomy & Cell Biology, University of Kansas Medical Center, Kansas City, KS; 2) Dept. of Biochemistry, Midwestern University, Downers Grove, IL.

Orofacial clefts are among the most frequent birth defects, affecting 1/800 births, in the U.S. alone. While a number of contributory genes have been identified, there is continued need to understand underlying pathogenetic mechanisms. Previously, we identified SPECC1L as the first gene mutated in a severe cleft that extends from the oral cavity to the eye, termed Oblique Facial Cleft (OFBC). Although less common, insights into cellular and molecular mechanisms underlying ObFC directly impact our understanding of more common facial malformations, including cleft lip. We have now created a mouse model of Specc1l deficiency using two independent genetrap alleles. Homozygous Specfll mutants are embryonic lethal with defects in delamination and migration of neural crest cells (NCCs). Cranial NCCs delaminate from early embryonic neural folds and migrate to the first and second branchial arches, which give rise to majority of craniofacial structures. Speclll is expressed in the neural folds at E8.5 (site of premaxillary NCC delamination) and in the branchial arches at E10.5 (destination of migratory NCCs). In rare cases, heterozygous Specfll mutant embryos regress frontal prominences late in development, consistent with the human phenotype. In vitro and in vivo cellular analyses indicate increased actin filaments and decreased AKT signaling upon SPECC1L deficiency. SPECC1L protein is stabilized at cell-cell boundaries upon confluence and interacts with both β-catenin and E-cadherin, two canonical components of adherens junctions associated with ECT1. Compared to WT E12.5 limb buds, RhoA activity is reduced 16-fold by reducing expression of SPECC1L. Thus, AKT is proposed to directly inhibit E-cadherin levels. However, reduced AKT signaling upon SPECC1L deficiency is also consistent with stronger AJs in kd cells. Regulation of cell-cell contacts is important not only for NCC delamination from the neural folds but also for migration of delaminated NCCs to their branchial arch destinations. We have confirmed altered β-catenin staining in epithelial and mesenchymal tissue in vivo. Thus, SPECC1L is an entirely novel modulator of AJ strength - the first to actually weaken AJs normally - and of AKT signaling affecting NCC function in facial morphogenesis.

Mutations in MAP3K1 tilt the balance from SOX9/FGF9 to WNT/β-catenin signaling. J.C. Loke, A. Pearlman, H. Ostrer. Pathology, Albert Einstein College of Medicine, Bronx, NY.

In-frame mis-sense and splicing mutations (resulting in a 2 amino acid insertion or a 34 amino acid deletion) dispersed through the MAP3K1 gene tilt the balance from the male to female sex-determining pathway, resulting in 46,XY disorder of sex development (DSD). These MAP3K1 mutations affect the balance by enhancing WNT/β-catenin/FOXO2 expression and β-catenin activity and by reducing SOX9/FGF9/FGFR2 expression. These effects are mediated at multiple levels involving MAP3K1 interaction with protein co-factors and phosphorylation of downstream targets. In primary lymphoblastoid cells and NT2/D1 cells transfected with wild type or mutant MAP3K1 cDNAs under control of the constitutive CMV promoter, these mutations increased binding of RHOA, MAP3K4, FRAT1 and AXIN1 and increased phosphorylation of p38 and ERK1/2. Overexpressing RHOA or reducing expression of MAP3K4 in NT2/D1 cells produced phenocopies of the MAP3K1 mutations. Reducing expression of RHOA or overexpressing MAP3K4 in NT2/D1 cells produced anti-phenocopies. Furthermore, the effects of the MAP3K1 mutations were rescued by co-transfection with wild type MAP3K4. Although MAP3K1 is not usually required for testis-determination, mutations in this gene can disrupt normal development through the function of genes demonstrated in this study.

Losartan increases bone mass by direct inhibition of osteoclasts. S. Chen, T. Sibai, N. Rianon, T. Yang, J. Black, E. Munivez, T. Bertin, B. Dawson, Y. Chen, B. Lee. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Boston University School of Medicine Orthopaedic Surgery, Boston University, Boston, MA, USA; 3) Department of Internal Medicine, University of Texas Medical School at Houston, TX, USA; 4) Laboratory of Skeletal Biology, Center for Skeletal Disease and Tumor Metastasis, Van Andel Research Institute, Grand Rapids, MI, USA; 5) Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, TN, USA; 6) Howard Hughes Medical Institute, Houston, TX, USA.

Osteoporosis and hypertension are two major chronic diseases of advanced age. Although traditionally these diseases are viewed as separate entities, increasing evidence suggests an overlapping etiology. One important determinant for both hypertension and osteoporosis is renin-angiotensin signaling. Inhibition of this signaling pathway lowers blood pressure and has demonstrated potential for bone loss prevention. We investigated the molecular mechanism underlying the regulation of bone mass by the renin-angiotensin pathway through animal studies and observational human data. Wild type mice were treated with Losartan, an antihypertensive drug that inhibits the angiotensin type 1 receptor, from birth until 6-weeks of age, after which bones were collected for microCT and histomorphometric analyses. Elderly hypertensive mice treated with Losartan were evaluated by Dual-energy X-ray absorptiometry and bone turnover markers at 6 and 12 months after the onset of treatment. Losartan increased trabecular bone volume vs. tissue volume (a 98% increase) and cortical thickness (a 9% increase) in 6-week old wild type mice. The bone changes were attributed to decreased osteoclastogenesis by decreasing RANKL expression per number per bone surface in vivo and suppressed osteoclast differentiation in vitro. At the molecular level, RANKL-induced ERK1/2 phosphorylation was attenuated by Losartan, suggesting a convergence of RANKL and angiotensin signaling at the level of ERK1/2 regulation. This also suggests that Losartan acts on ERK1/2, the essential mediator of RANKL signaling, and influences osteoclast differentiation. Two women (Age 81 and 74) were prospectively treated for idiopathic hypertension with Losartan. Interestingly, both showed evidence of decreased bone resorption as measured by the urinary N-telopeptide. Altogether, inhibition of the angiotensin pathway by Losartan has beneficial effects on bone beyond reducing blood pressure. Our study adds evidence to support the relationship between angiotensin receptor blocker benefits (Losartan) and age-related bone loss that increases the risk of fractures in the elderly.

The role of the ELOVL gene family in neurodevelopmental disorders. J. Gerard, A. Moreno-Di-Luca, D.W. Evans. 1) Neuroscience Department, Bucknell University, 701 Moore Avenue, Lewisburg, PA 17837; 2) Geisinger-Bucknell Autism and Developmental Medicine Center, 120 Hammad Drive, Lewisburg, PA 17837.

Very long-chain fatty acids (VLCFAs) are essential for basic cell structure and multiple cellular functions including intercellular signaling. Fatty acid elongases, encoded by the elongation of very-long-chain fatty acids (ELOVL) gene family, catalyze the first and rate-limiting step in VLCFA synthesis, condensing acyl-CoA and manoyl-CoA to produce 3-ketoacyl-CoA. Mammals have seven elongases (ELOVL1-7), each of which exhibits a characteristic pattern of expression and substrate specificity. Previous research suggests that abnormalities in the ELOVL genes may cause a wide range of disorders. Heterozygous mutations in ELOVL4 result in autosomal dominant Stargardt macular dystrophy 3, whereas homozygous mutations cause a severe neurodevelopmental disorder characterized by ichthyosis, spastic quadriplegia, and intellectual disability. Elovl3-null mice display a tissuedevelopmental delay, increased bodyweight, increased bone mass, and increased transgenic mice present with obesity, hepatic steatosis, and insulin resistance. ELOVL7 overexpression is thought to play a role in prostate cancer growth. By searching through the International Standards for Cytogenomic Arrays (ISCA) Consortium database, a repository of whole genome chromosomal micro array (CMA) data from patients with clinically diagnosed neurodevelopmental disorders, we identified two individuals with overlapping deletions in chromosome 5q12.1 with a smallest region of overlap including a single gene, ELOVL7. Case 1 was referred for autism and developmental delay and found to have a 135 kb deletion involving only ELOVL7. Case 2 presented with oromotor apraxia and developmental delay and has a 275 kb deletion spanning ELOVL7, DEPDC1B (DEP domain containing 1B), and ERC6 (excision repair complementing rodent mutation 6). Our data support the idea that mutations in ELOVL7 contribute to the human phenotype and may play an important role in the development of congenital anomalies.
3184W
AKT1 gene mutation levels are correlated with the type of dermatologic lesions in patients with Proteus syndrome. M.J. Lindhurst1, J. Wang2, H. Bloomhardt2, A.M. Wilkowskii1, L.N. Singh1, D.P. Bick2, M.J. Gambello3, C.M. Powell3, C.R. Lee4, T.N. Darling4, L.G. Biesecker1. 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Department of Dermatology, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; 3) Department of Pediatrics and Obstetrics & Gynecology, Medical College of Wisconsin, Milwaukee, WI, USA; 4) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 5) Departments of Pediatrics and Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 6) Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MD.

Proteus syndrome (PS) is characterized by progressive, mosaic, segmental overgrowth and occurs sporadically. PS is caused by a post-zygotic somatic activating mutation c.49G>A, p.Glu17Lys in AKT1. To date, all patients who meet the clinical diagnostic criteria for PS and have been tested in our laboratory have this mutation. Skeletal overgrowth and dermatologic lesions are the most common manifestations of PS. Cerebriform connective tissue nevi (CCTN) are a highly specific and common lesion in patients with PS and are characterized by a massively expanded dermis filled with thick collagen bundles. Epidermal nevi (EN) can occur sporadically or as part of several syndromes including PS. They have a rough surface, are dark in color, usually follow Blaschko lines of Blaschko, and exhibit hyperkeratosis, papillomatosis, and acanthosis. It is unknown which cells determine the formation of these lesions. Based on the histology, we hypothesized that CCTN were generated by mutation positive cells in the dermis and that EN were generated by mutation positive cells in the epidermis. To test this hypothesis, we isolated fibroblasts and keratinocytes from CCTN, EN and apparently normal skin samples and measured the level of the mutant allele in each cell type. The mutation level in the fibroblasts isolated from seven CCTN biopsies from the feet of three patients ranged from 9-32%. The mutant allele was not found in any of the keratinocyte cultures from these samples. In four biopsies from apparently unaffected skin in two patients, the mutation level in the fibroblasts was 6-27% and there was no evidence of the mutant allele in the keratinocytes. The mutation level in fibroblast samples from normal skin was measured in 10 samples and ranged from 0-38%, whereas in keratinocytes isolated from these EN, the mutation level was 0-44%. We conclude that the AKT1 p.Glu17Lys activating mutation in keratinocytes is a key determinant of EN formation. The inability to detect the mutates could be due to the EN keratinocyte cultures being derived from normal skin.

Function of miR-199a-5p in Stage-specific Osteogenesis of Human Mesenchymal Stem Cells. S. Gu1, X. Chen1, B.F. Chen1, G. Li1, H.W. Ouysang2, Y. Wan1, T.L. Lee1, W. Chen1. 1) Institute of Biomedical Sciences Building, The Chinese University of Hong Kong, HKSAR, Hong Kong; 2) Zhejiang Provincial Key Laboratory of Tissue Engineering and Regenerative Medicine, Zhejiang University, Hangzhou, Zhejiang, China.

Elucidating the regulating mechanisms of osteogenesis of human mesenchymal stem cell (hMSC) is important for the development of cell therapies for bone loss and regeneration. MSC differentiation involves complex pathways that are regulated at both transcriptional and posttranscriptional levels. However, the key regulator(s) of MSC differentiation has not been identified. microRNAs (miRNAs) have been shown to regulate almost every biological process, including stem cell differentiation. Here, we show that miR-199a-5p modulates osteogenic differentiation of hMSCs. miR-199a-5p expression detected by qPCR revealed its up-regulation during osteogenesis of hMSCs. Over-expression of miR-199a-5p but not -3p enhanced differentiation of hMSCs both in vitro and in vivo, whereas inhibition of miR-199a-5p by miR-199a-5p siRNA reduced osteogenesis of hMSCs. In order to study the underlying mechanism, we generated an miR-199a-5p overexpression cell line in which the expression was driven by a hypoxia inducible factor (HIF) responsive promoter. In this cell line, miR-199a-5p expression is up-regulated by HIF1α activation. Constitutive expression of miR-199a-5p enhanced osteogenesis maturation by inhibiting the expression of key osteogenic regulator genes. To test whether miR-199a-5p regulated the expression of osteogenic regulators at transcriptional and posttranscriptional levels, we performed qPCR and western blot analysis and found that miR-199a-5p regulated the expression of early and late osteogenic regulators at both transcriptional and posttranscriptional levels. In conclusion, our findings suggest that miR-199a-5p may be a negative regulator of the osteogenic process and its function in osteogenesis may be context-dependent.

3186W
DFLAT: Functional Annotation for Human Development, H.C. Wick1, D.P. Hill2, H. Drabkin2, H. Ngu3, M. Sackman1, C. Coumier1,2, J. Haggett4, J.A. Blake2, D.W. Bianchi2, D.K. Stomn1,2. 1) Department of Computer Science, Tufts University, Medford, MA; 2) Bioinformatics and Computational Biology, The Jackson Laboratory, Bar Harbor, ME; 3) Tufts University School of Medicine, Boston, MA; 4) Mother Infant Research Institute, Tufts Medical Center, Boston, MA.

Recent technological advances have led to a notable increase in genomic studies of the developing human fetus and neonate. Interpreting these studies requires widespread characterization of the functional roles of genes in different organs at appropriate developmental stages. The Gene Ontology (GO), a valuable and widely-used resource for characterizing gene function, offers perhaps the most suitable functional annotation system for this purpose. However, due to our effort in studying molecular genetic effects in humans, even the current collection of comprehensive GO annotations for human genes and gene products is often inadequate for scientists wishing to study gene function during human fetal development. The Developmental Functional Annotation at Tufts (DFLAT) project aims to improve the quality of analyses of fetal gene expression and regulation by curating human fetal gene functions using both manual and semi-automated GO procedures. Eligible annotations are then contributed to the GO database and included in GO releases of human data. In addition to augmenting GO directly, DFLAT has produced a considerable body of functional annotation that, although too preliminary for incorporation in GO, may provide valuable information about developmental genomics. This includes developmentally relevant annotations transferred from mouse orthologs to human genes. A collection of gene sets combining existing GO annotations with the DFLAT annotations is now available in the Gene sets browser.

The results demonstrate widespread changes with DFLAT and an increase in the number of implicated gene sets. Blinded literature review supports the validity of newly significant findings obtained with the DFLAT annotations. New insights and significant gene sets suggest potential hypotheses for future research. Overall, the DFLAT project contributes new functional annotation and gene sets likely to enhance our ability to interpret genomic studies of human fetal and neonatal development.

3187W
Transcriptome and pathway analysis of fetal and adult human retina, RPE, and choroid. A.S. Boleda1, M. Brooks1, A. Maminishkis1, S. Miller1, A. Swaroop2. 1) National Eye Institute, National Institutes of Health, Bethesda, MD. 2) Departments of Neurology and Ophthalmology, Medical College of Wisconsin, Milwaukee, WI, USA; 4) Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 5) Departments of Pediatrics and Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 6) Laboratory of Pathology, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Current knowledge of the human retina, retinal pigment epithelium (RPE), and choroid is derived from a variety of in vitro and in vivo models. Thus, functional differentiation of human tissue transcriptomes was carried out using RNA-Seq in fetal and adult tissues. By determining the expression of genes differentially associated with adult and developmental retinas, we build a foundation of knowledge for disease investigation. In this study, RNA-Seq was performed for human fetal samples (18-19th week post-conception) of three ocular tissue types (retina, RPE, and choroid) as well as 3 adult foveal and peripheral retinal tissues (86, 87, and 92-year-old healthy eyes). Directional RNA-seq libraries (27 fetal and 6 adult) were prepared following a modified version of the standard Illumina TruSeq RNA protocol. Illumina GAIIx technology was employed for single-end sequencing to 76 nucleotides. Paraffin-embedded tissue sections were stained with antibodies to marker proteins for the following: EMA, Iba1, GFAP, and Brn3a, and counterstained with hematoxylin. Digital image analysis was performed using GenePaint software and the Tuxedo Suite. The JMP 10 package was implemented for computing principal components (PCA), creating hierarchical clusters and filtering lists of differentially expressed genes. Further, pathway and network analysis was performed using GOrilla. Approximately 32 million reads per sample mapped uniquely to the Hg19 reference genome. About 28,000 transcripts were expressed at an RPKM value greater than 1.0 in any tissue, and of these between 10-15,000 transcripts showed differential expression between tissue types. PCA demonstrated that samples were grouped by tissue type, and initial pathway analysis was in concordance with our predictions. Three sets of differential analysis comparisons were made. The first compared the fetal tissue types, the second compared the adult and fetal retinas, and finally we compared the adult tissue types. A number of up-regulated genes in the fetal retina were involved in neuronal differentiation, synaptic transmission and visual perception pathways. Those up-regulated in the fetal RPE are important to pigment biosynthetic processes, vitamin A metabolic processes, and light detection. Finally, up-regulated genes in fetal choroid were involved in development for anatomical and structural morphogenesis, cellular developmental processes, response to stimuli, and multiple pathways important to immune response.
RING CHROMOSOMES ABERRATIONS AT A PEDIATRIC MEXICAN HOSPITAL. TWO CASES WITH MOSAISM OF CHROMOSOME 13, 46XY / 46, XY, r (13) AND CHROMOSOME 18, 46, XY / 46, XY, r (18). M. Hurtado-Hernandez1, J.M. Aparicio-Rodriguez2, M. Barrientos-Perez2, S. Chatelain-Mercado2. 1) Cytogenetics; 2) Genetics; 3) Endocrinology, Hospital para el Nino Poblano, Puebla, Puebla; 4) Estomatología, Benemerita Universidad de Puebla; 5) BioTeCnology, Universidad Autonoma Metropolitana, Mexico.

The autosomic alteration due to a ring formation is a rare aberration of either chromosome 13 and 18 which is in relation with phenotypic malformations, neurologic problems and genital abnormalities. Two clinical polymorphic cases with skull treboliform dismorphies with early seizures and malformed genitals with microopenis is presented from four of the total patients found in this study. Among chromosomal alterations, the ring of autosomic chromosome 13 and 18 are not frequent, the main phenotypical alterations in this study are in relation to neurological, genital and craniofacial malformations. Taking in consideration that mutations or chromosome aberrations are alterations in the chromosome number or structure. They are mainly considered due to gametogenesis inborn error (meiosis) or during the zygote first cellular divisions. All these alterations might be observed during metaphase from the cellular cycle, where DNA losses are seen due to DNA repair processes deficiency o total absence, among others. 4617 chromosomal studies were performed at Hospital Para El Nino Poblano (Pediatric Hospital) in Mexico (from 1992 to 2011) were 34.6% (1596 patients) showed different chromosomal alterations and only two patients showed ring chromosome aberrations. These chromosome changes are classified as structural alterations. Both pediatric patients with these genetic diseases are described in this study analyzing their clinical characteristics and medical or surgical treatments according to the phenotypic alterations.

Inverted Duplication with Terminal Deletions: Variations on a Theme. N. Christacou1, S. Schonberg1, P. Mowrey2, 1) Dept Cytogenetics, Quest Diagnostics, Chantilly, VA; 2) Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD.

Chromosomal rearrangements resulting in an interstitial inverted duplication with associated terminal deletion, first described in 1976, have become well-established as recurrent underlying causes of genomic imbalance in man, and have been described in association with multiple chromosome regions. The application of newer molecular cytogenetic methods, in particular microarray, has allowed for a significantly higher level of resolution for detecting and characterizing these and other chromosomal abnormalities. Three mechanisms are commonly proposed to explain the origin of inverted duplications with terminal deletions, all involving formation of a dicentric chromosome intermediate that subsequently breaks during meiosis to form a monocentric duplicated and deleted chromosome. Reported cases in the literature nearly all follow a pattern of a relatively distal deletion breakpoint with end stabilization at breakpoint associated repeat sequences, likely due to ascertaining bias as others would typically result in a lethal phenotype. Here we report the cytogenetic results of three cases that do not follow this typical pattern and may not be completely explained by previously proposed mechanisms. Case 1 involves a derivative chromosome 14 with a proximal 14q breakpoint secondary to inverted duplication with terminal deletion and capture of the deleted region of chromosome 14 by the 7p telomeric region. Inverted duplication of chromosome 14 (14q11.2q13.1) was confirmed with FISH using the T-cell receptor LSI TRA/D probe (Abbott Molecular). As the entire deleted region was captured intact, a dicentric intermediate may not fully explain this result. Case 2 involves a derivative chromosome 15, also with a proximal long arm breakpoint secondary to inverted duplication with terminal deletion and fetal rescue by chromosome 15 nordsisjunction. Case 3 has the commonly encountered inverted duplication of 8p (8p11.2p23.1) in which the inv dup (8p) chromosome was stabilized at its breakpoint by telomere capture of distal 4p resulting in duplication bands 4p15.2 to the 4p terminus. These cases illustrate how the use of newer molecular cytogenetic technologies are crucial to identifying these uncommon variations of interstitial duplication with associated terminal deletion and present a cautionary note that many such rearrangements may not follow the most common pattern described in the literature.

Infertility related to a rearrangement of the Y chromosome. P.A. Delgado1, S. Iyer2, A. Jarin1, N. Rao1, C.A. Tirado1. 1) UCLA Dept. of Pathology and Laboratory Medicine, Los Angeles, CA; 2) UT Southwestern Medical Center Dallas, Texas.

Structural abnormalities of the Y chromosome are the most frequent chromosomal aberrations in infertile male. Herein, we present a 30 year-old male with a history of infertility. A physical examination reveals no characteristics associated with any gonosomal aneuploidy. Further studies also showed oligospermia. Chromosome analysis of peripheral blood shows what seems to be a deletion of the Y chromosome first described as 46, X, del(Y)(q11.23). C-banding was negative for the heterochromatic region normally found on the distal long arm (q) of the Y chromosome. It couldn’t be determined by this analysis if part of the euchromatic band q11.23 has been deleted. FISH analysis using the SRY probe showed two signals for the SRY gene. Molecular genetic studies showed that there was no deletion of the AZ regions. This patient had a derivative (der) Y chromosome which results in partial monosomy of the long (q) arm and partial disomy for the short (p) arm. The ISCN was modified as: 46,X,der(Y)(pter->q11.23::p11.2->pter);ish der(Y)(DYZ3+;SRY++). Most of these cases are mosaic. Structural rearrangements of the Y chromosome are very common. Their influence on gonadal and somatic development is extremely variable depending on the genomic sequences duplicated or deleted and the variable degrees of mosaicism. However, rearrangements of the Y chromosome will often show a clinical picture of azospermia or oligospermia. Further molecular genetic testing for microdeletions of the long arm of the Y chromosome was recommended. Genetic evaluation of the patient was also suggested.

AFurther Case of de novo Isochromosome 18q with additional skeletal system abnormalities. E. Karaca, T.R. Ozdemir, A. Durmaz, F. Ozkinay, O. Cogulu. Ege University Faculty of Medicine, Department of Medical Genetics, Izmir, Turkey.

Isochromosome 18q is a rare chromosomal disorder presenting a variety of phenotype ranging from mild facial dysmorphism to severe malformations. Findings in isochromosome 18q overlap with isochromosome 18p and trisomy 18, and the clinical picture has been reported from feotuses. Because very few cases have been described in the literature, the phenotypical features have not been fully well-defined because of the rarity of those cases. Main features can be summarized as holoprosencephaly, heart defects, defects of the gastrointestinal and genitourinary systems and extremity anomalies. Here, we present a new case of isochromosome 18q. The proband is a 16 months old female who was born to nonconsangunous parents. She was first admitted to the hospital at 15 days old because of pulmonary insufficiency and has been hospitalized twice because of pulmonary infections and developmental delay. On admission her weight was 4880 g (3p), height: 53 cm (<3p), and her head circumference was 38 cm (25p). Her physical examination showed short stature, exophtalmos, right eye ptosis, blue sclera, low-set ears, long philtrum, high palate, micrognathia, pectus excavatum, short extremities, fusiform fingers, pes equino-valgus, hallux valgus and club foot. Immunological and biochemical laboratory analysis were normal. Echocardiography revealed ventricular septal defect. Visual electrophysiology showed bilateral partial conduction defects. Abdominal and cranial USG were normal. She had sensorineural hearing loss (right ear 60dB, left ear 75dB). Cranial MRI showed corpus callosum hypoplasia and dilated anterior subarachnoid space. Bone survey revealed distinctive skeletal system abnormalities such as small jaw, pectus excavatum, coxa valga, delayed ossification of hip and bilateral pes equinovarus. Her karyotype was 46,XX, i(18)(q10). Her parents’ karyotypes were normal. Skeletal system abnormalities such as pectus excavatum and coxa valga were described in trisomy 18 but not in isochromosome 18q syndrome. In conclusion our case contributes to the literature in two ways; first additional skeletal anomalies, and second reporting a further case of a very rare chromosomal disorder.
3192T

Tetrasomy 13q32.2qter Due to an Apparent Inverted Duplicated Neo-centric Marker Chromosome in an Infant With Hemangiomas, Failure to Thrive, Laryngomalacia, and Tethered Cord. J. Liu1, M. Del Vecchio2, D. Pezanowski3, H. Punnett4, J.P. de Chadarévian5. 1) Dept Path & Lab Med., St Chris Hosp for Children, Drexel University College of Medicine, Philadelphia, PA 19134; 2) Department of Pediatrics, Temple University School of Medicine, Philadelphia, PA 19140.

Fewer than 100 small supernumerary marker chromosomes (sSMCs) with a non-α-satellite neocentromere structure have been reported in the literature. The major morphology of a neocentric sSMC is an inverted duplication of a distal chromosome end, while ring chromosomes and centric minutes represent a small percentage. Fourteen percent were derived from segments on 13q with four breakpoints (13q14, 13q21, 13q31, 13q32) described. To date, only 5 neocentric sSMCs containing inverted duplicated segment 13q32qter have been reported, and three of the 5 cases were deceased when the manuscripts were published. Here we report a 9 week old African American male who was admitted with laryngomalacia and failure to thrive. He had multiple hemagiomas on the tip of the nose, the mid-forehead, the scalp, and the sacrum. Other abnormalities observed included micrognathia, posterior rotation of the ears, sacral dimples, tag, and cleft, tethered cord, mild splenomegaly, elevated hemoglobin/hematocrit and thrombocytopenia. He was a 34-week preemie with a normal newborn screen and hearing test. Family history was negative, consanguinity was denied. A month after the hospital visit, the proband was found dead in his crib. Chromosomal Microarray Analysis (CMA) with Affymetrix’s CytoScan™ HD SNP array revealed a terminal triplication on the long arm of chromosome 13 from 13q32.2 to 13qter (tetrasomy 13q32.2qter), spanning approximately 15.6 Mb in size and encompassing 16,495 markers/probes. Reflex chromosome analysis demonstrated a small supernumerary marker chromosome (sSMC) in 95 of 100 metaphases examined that was consistent with a high percentage (95%) mosaicism karyogram. The morphology of the identified sSMC was suggestive of an inverted duplication (inv dup) of a chromosomal segment and, based on CMA findings, most likely comprises two extra copies of the segment 13q32.2qter. Reflex FISH study was consistent with the structure of an inverted duplication with a neocentromere. Genotype pattern of the involved SNPs by CMA suggests a L deletion exchange at melano-1. Parental cytogenetic studies were negative. We have described in this report the sixth case of an apparent neocentric sSMC containing an inverted duplicated chromosomal fragment 13q32qter characterized by genome wide SNP array, conventional cytogenetics and FISH studies. The documented clinical details will assist the medical genetics community to better understand this rare chromosomal disorder.

3193F

Prenatal diagnosis of a complex 9 break rearrangement requires karyotype, microarray and whole-genome sequencing. M.J. Macera1, A. Sobrino1, B. Levy2, V. Jebanputra3, V. Apparaval2, A. Mills1, C. Esteves2, C-Y. Yu3, C. Hanscom4, V. Pillaiamani4, M. Talkowski4, D. Warburton2. 1) Department of Genetics, New York-Presbyterian Hospital, Columbia University Medical Center, New York, NY; 2) Department of Pathology and; 3) Genetics and Development and Pediatrics, Columbia University, New York, NY; 4) Center for Human Genetic Research, Massachusetts General Hospital, Department of Neurology, Harvard Medical School, Boston, MA.

Complex chromosomal translocations, defined as apparently balanced constitutional structural rearrangements involving three or more chromosomes or more than two breakpoints, are rarely detected in prenatal testing. Only 0.03% from a survey of 269371 prenatal studies were determined to be de novo (Giardino et. Al. 2009). We report a case with 5 chromosomes involved in a four way translocation and a separate two way translocation; both arms of the same chromosome 18 were involved in separate translocations. A 28 year old primi-gravida woman presented for amniocentesis sampling at 21 weeks gestation. Ultrasound and MRI revealed bilateral ventriculomegaly (13mm and 15mm) and colpocephaly, with partial agenesis of the corpus callosum. Her prior family history was unremarkable with no unusual environmental exposure. Cytogenetic and FISH analysis with telomere probes on amniocytes revealed a 46,XX,i(3;18;5;7)(p25;p11.2;q13.3;q32), t(9;18)(p22;q21) karyotype in all cells examined. SNP oligonucleotide microarray analysis (SOMA) on fetal DNA showed no loss or gain of chromosomal material at any breakpoints. The pregnancy was terminated because of ultrasound findings. The unusual complex karyotype was confirmed in fetal kidney cells. Both parents had normal chromosomes. Next generation sequencing of fetal genomic DNA using large-insert jumping libraries, followed by PCR and Sanger validation identified minimal chromosomal losses and or gains. However 7 OMIM annotated genes were disrupted at the breakpoints. CNTN6 and TBC1D5 are interrupted on chromosome 3, CNTNAP2 on chromosome 7, PTPRD on chromosome 9 and L3MBTL4, LOC100130480 and WDR7 on chromosome 18. The chromosome 5 breakpoint did not involve any gene disruption. Sequencing revealed even more complexity. It was determined that small portions of chromosomes 3 and 7 were inserted into the chromosome 5 breakpoint and the p arm of chromosome 18 has a 184.5 kb inversion at the chr3/chr18 junction. This brings the total number of breaks in this chromosome complement to 9. The characterization of this extremely complex abnormality illustrates the necessity of both cytogenetic and molecular testing. G-banding coupled with telomere and painting probes detected the initial rearrangements. While microarray analysis showed no pathogenic gain or loss of material at the breakpoints of the translocations, sequencing found 7 genes disrupted by this rearrangement, as well as an even more complex chromosomal rearrangement.
Familial translocation t(4;8p) associated with a phenotype of combined hyperlipidemia. N. Quaresemin1, N. M. L. M. Castro2, C. H. P. Grangeiro1, J. A. Joshiakhan3, C. M. Loureiro3, L. A. F. Laureano1, J. H. Hube1,2, E. S. Ramos1,2, L. Martelli2,3. 1) Serviço de Genética Médica, Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto-USP, Ribeirão Preto, São Paulo, Brazil; 2) Laboratório de Citogenética - Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - Ribeirão Preto - SP, Brazil; 3) Departamento de Genética - Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brazil.

Combined hyperlipidemia affects around 2% of individuals and its etiology remains unknown. Lipoprotein lipase gene is located on short arm of chromosome 8. We report a case of a 17-year-old girl with short stature, dysmorphic features and hypertriglyceridemia associated with a familial translocation t(4;8p). The phenotypic abnormalities associated with this translocation include combined hyperlipidemia, with a gain in the 6p12.2-q12 region, with a size of 15,082 Mb, was performed for combined hyperlipidemia. We report four members of a family with rearrangements between chromosomes 4 and 8, two apparently balanced translocations and two derivatives of chromosome 4. The first patient (P1) was referred to Medical Genetics Division due to seizures and microcephaly. Chromosomal analysis by GTG banding showed a karyotype 46,XX,der(4;1-4;8)(p16;p23.1). The paternal karyotype was 46,XY, t(4;8)(p16;p23.1) and maternal karyotype was normal. Her phenotypic abnormalities including psychomotor retardation, low birth weight and genital anomalies as in the case of this patient, as well as craniofacial, brain and limbs abnormalities, microcephaly and cardiac abnormalities. It is proposed that the phenotypic presentation may be due to the complexity of the inversion mechanisms, since there have been only a few studies describing full clinical features of the patients with this translocation.

Clinical and molecular characterization of a subtelomeric deletion of 19p13.3 including STK71 with Peutz-Jeghers syndrome. S. Ishimaru1, N. Kurita2, T. Murakoshi2, R. Fukuzawa3, T. Kuchikata4, H. Yoshishashi5. 1) Division of Hematology and Oncology, Tokyo Metropolitan Children’s Medical Center, Tokyo, Japan; 2) Division of Gastroenterology, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 3) Division of Pathology, Tokyo Metropolitan Children’s Medical Center, Tokyo, Japan; 4) Division of Medical Genetics, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan.

The subjects with a deletion of 19p have been infrequently reported. Additionally, there have been only a few studies describing full clinical features of the patients with this deletion. In our series, one patient with a large deletion of 19p13.3 was evaluated. Peutz-Jeghers syndrome (PJS,MIM175200) is an autosomal dominant disorder characterized by pigmented macules on the lips and oral mucosa and multiple gastrointestinal hamartomatous polyps. The phenotype of this disorder is highly variable and can include other diseases such as breast cancer, thyroid cancer, and peptic ulcer disease. However, knowing the characteristic sequence variants in STK71 (serine-threonine kinase 11) located on 19p13.3, some are due to exons and whole-gene deletions.

Association between abnormal phenotype and chromosome heteromorphisms. J. Grzesiuk1, C. S. Pereira2, F. G. O. Gennaro3, L. A. F. Laureano2, S. A. Santos1, J. Huber4, L. Martelli2,3,1. 1) Dept of Genetics, University of São Paulo, Ribeirão Preto, SP, Brazil; 2) Clinical Hospital, School of Medicine, Ribeirão Preto, SP, Brazil.

Heteromorphisms are variations among individuals in heterochromatin regions such as inversions, duplications and deletions and they are considered benign changes. However, literature suggests its association with several sporadic phenotypes, including Walker-Warburg syndrome. This study aimed to describe the frequency of different chromosome heteromorphisms in patients referred for diagnostic investigation and the phenotype of the carriers. We have reviewed GTG banding karyotypes of 2,892 patients referred to the Medical Genetics Division at the Clinical Hospital (HCPR) between 2006 and 2013. Among them, 92 patients (3.18%) presented heteromorphic chromosomes, similar to the frequency of 2-5% observed by Gardner and Sutherland (2004) in the general population. This finding corroborates with the literature suggesting that heteromorphisms are not involved with phenotypic changes. However, among the heteromorphic patients, other phenotypic findings, as mental retardation and facial dysmorphism, were described in 61 patients (66.30%), four times more frequent than the reproductive disorders phenotype diagnosed in 16 patients (17.35%). Increase in length of the heterochromatin on the long arm of chromosomes 1, 9, 16 and Y represented 29.35% of the heteromorphisms, while the satellites and stalks on the short arm of chromosomes 14, 15, 21 and 22 were detected in 16.30% and chromosome 9 inversions were seen in 34.78%. We have also noticed a high rate of co-occurrence of heteromorphisms, 18 of the 92 heteromorphic patients (19.57%) presented two heteromorphisms or one heteromorphism concomitant to other type of chromosome abnormality, that reinforces the hypothesis of the existence of a chromosome interaction (2011). Some variants were located on chromosome 9 and the pericentric inversion of chromosome 9 was the main finding associated with positive phenotypes. The differences in the size of the segments, the occurrence of two chromosomal breaks and the involvement in the chromosome interaction (2011). The latter was related to the occurrence of small losses or gains of critical genetic material and consequent more severe phenotypes. Genomic studies and gene expression analysis should be used for characterization of the heteromorphisms and better understanding about its phenotypic effects.
3198T
Ring chromosome 11: Familial case with normal development and short stature. Further delineation of this rare cytogenetic abnormality. A. Singer1, R. Berger2, R. Segal3, C. Vinkler4. 1) Dept Clinical Gen Unit, Barzilai Med Ctr, Tel Aviv, Israel; 2) Cytogenetic Lab, Maccabi Health Services Rehovot Israel; 3) Institute of Medical Genetics, Shaare Zedek Medical Center Jerusalem, Israel; 4) Institute of Medical Genetics, Wolfson Medical Center, Holon, Israel.

Autosomal ring chromosomes are uncommon cytogenetic aberrations identified in prenatal and postnatal diagnosis. It is estimated to occur in less than 1:30,000. Rings have been reported for all chromosomes, although those involving autosomes 13 and 18 are among the most common. The classic mechanism of ring formation is breakage in both arms of a chromosome followed by fusion of the two broken arms and loss of the distal segments. Therefore, phenotypic abnormalities associated with partial deletions, can be found among patients with ring chromosomes. Only a few examples of parental transmission are known and it has been estimated that about 99% of all ring chromosomes arise sporadically. Ring chromosome 11 is rarely observed and only a few cases have been reported, almost all have growth failure, variable phenotypes and some degree of intellectual disability. We present a case of a mother and daughter with ring chromosome 11. The girl is 8 years old and was referred to our clinic for evaluation because of short stature. Her parents are healthy and nonconsanguineous. The father’s height is normal and mother’s height is 153 cm (3rd centile). The girl’s height is 112 cm (-2.5 SD), she has three hypopigmented skin lesions, myopia, no significant dysmorphic features and normal development. Cardiac echo was normal. GTG-banded chromosome analysis on peripheral blood of the proband showed a karyotype 46, XX, r(11)(q15;p25) in all metaphases. Paternal and maternal chromosomes showed equal mosaic of 45XX/11q-, 46XX,r(11)(q15;p25). This is the first description of familial ring chromosome 11 with short stature and normal development lacking any other phenotypic abnormality. The phenotypic differences in previously reported cases are most likely due to the differences in the size of the terminal deletion(s) at one or both arms or the result of mitotic instability. Further evaluation is being done with SNP array in order to define the deleted area addressing the mild phenotype in our family. Detailed molecular characterization and documentation of the deleted region is ongoing.

3199F
Mosaic Turner Syndrome with unilateral absence of digital phalanges and renal agenesis. h. ulucan, a. koparir, g. guven, a. celebi, e. koparir, m. seven, m. ozen. Department of Medical Genetics, Istanbul University Cerrahpasa Medical School, Istanbul, Turkey.

Turner’s syndrome (TS) is a sex chromosome disorder occurring in 1 in 2,500 female newborns and in approximately 50 in 100,000 adult females and is due to partial or total loss of the second sex chromosome. There is great variability in cytogenetic findings, including the 45, X karyotype, mosaics without structural abnormalities (as mos 45,X/46,XX) and mos 45,X/46,XY) and structural abnormalities with or without mosaicism (as isochromosomes and marker chromosomes). The 45,X/46,XX chromosomal pattern is the most frequent mosaic type of this disease (36%). TS is characterized by retarded growth, gonadal dysgenesis and infertility. Renal and/or collecting system malformations have been found in 30-40% of cases; including horseshoe kidney, renal malrotation and collecting system malformations. Skeletal anomalies are frequent but severe hand anomalies were not reported. Here, we report five-year-old female with unilateral absence of digital phalanges and unilateral renal agenesis whose chromosomal analysis showed 45X,46XX. TS is known to cause multisystemic disorders. Here we present a case with unilateral renal kidney which is reported as a rare anomaly. Skeletal anomalies are frequent but severe hand anomalies were not reported. Here, we report five-year-old female with unilateral absence of digital phalanges and unilateral renal agenesis whose chromosomal analysis showed 45X,46XX. TS is known to cause multisystemic disorders. Here we present a case with unilateral renal kidney which is reported as a rare anomaly.
Duplications of chromosome Xq28 including the MECP2 gene have been described primarily in male patients with severe developmental delay, spasticity, epilepsy, stereotyped hand movements and recurrent infections. Females carriers are most of the time asymptomatic and display an extreme skewing of chromosome X-inactivation. Symptomatic females have been reported in 15 cases in the literature, the majority with a milder phenotype. The duplications resulted in functional Xq disomy or were de novo or inherited intrachromosomal Xq duplications associated with a random X inactivation. The duplications were maternally inherited in only 3 cases, the others were de novo. We carried out in 2012 a national study that permitted to identify in France 11 affected females, including 2 twins, aged 4.5 to 40 years, among the 100 patients carrying a MECP2 duplication. Five were intrachromosomal duplications of Xq, 3 resulted from an unbalanced X-autosome translocation, 1 from a de novo duplication resulting in functional Xq disomy or were de novo or inherited intrachromosomal Xq duplications associated with a random X inactivation. The duplications were maternally inherited in only 3 cases, the others were de novo. We carried out in 2012 a national study that permitted to identify in France 11 affected females, including 2 twins, aged 4.5 to 40 years, among the 100 patients carrying a MECP2 duplication. Five were intrachromosomal duplications of Xq, 3 resulted from an unbalanced X-autosome translocation, and 2 were small size intragenic MECP2 duplications, which causality determination is on going. The size of the duplications ranged from 5.5 Kb to 11.7 Mb. The duplications were de novo in all 7 patients with available parents. Together with the literature, we showed that females with X-autosome translocations display a severe phenotype similar to the male ones, while others display an unspecific phenotype. In our study 22% of affected female patients presented with spasticity, 33% with epilepsy, 66% with stereotyped hand movements, and 56% with recurrent infections. These results are of importance for genetic counselling since an abnormal phenotype in females born from carrier mothers is very rare.

**Syndromic intellectual disability in a patient with 3.5 Mb deletion at 1p13.3**

We report the case of a 53-year-old man with intellectual disability, short stature, and frequent hospitalisations due to recurrent infections. The genetic analysis revealed a terminal deletion of the long arm of chromosome 1 of 3.5 Mb that included the MECP2 gene. The patient has been followed up for over 10 years and has shown mental and physical regression. The deletion was confirmed by FISH analysis and detailed molecular cytogenetics. The presence of the MECP2 gene in the deletion region is likely to have contributed to the patient's clinical features.

**Clinical features in a pediatric population due to chromosome deletions at a third level pediatric Mexican hospital in 19 years period of time**

Five case reports. S. Chatelet1, I.M. Aparicio-Rodriguez1, M.L. Hurtado-Hernandez1, M. Barrientos-Perez1, P. Zamudio-Meneses1, M. Palma-Guzmán1, H. Chavez-Ozeki1. 1) Biotechnology, Universidad Autónoma Metropolitana, Mexico City; 2) Genetics; 3) Cytogenetics; 4) Endocrinology; 5) Cardiology, Hospital Para El Niño Poblan Boíl Del Nino Poblan; 6) Estomatología, Benemérita Universidad Autónoma de Puebla, Mexico.

Chromosome aberrations are considered as alterations in the chromosome number or structure. They are mainly due to gametogenesis inborn error (meiosis) or occur during the zygote first cellular divisions where DNA repair processes are deficient. Two Wolf Syndrome patients and two cri-du-chat patients with deletion of the short arm of chromosome 4 and 5, respectively, and deletion of chromosome 8 in one patient, were observed among 4107 karyotype studies performed from 1992 to 2011, at the Hospital Para El Niño Poblan (Pediatic Hospital) in Mexico. These chromosome structural alterations or deletions at chromosomes 4, 5 and 9, observed among five patients from three different families were chosen to analyze their clinical characteristics, medical and surgical treatments and their medical evolution according to the genetic disease.

**Molecular cytogenetic analysis of 8p23.1: about two Tunisian patients.**

H. Elhezali1, M. Kammoun1, L. Ben Haj Hamida2, H. Hannouch3, N. Soyah3, E. Elhezali1, A. Saad4, S. Mououg-Zrelli3. 1) Laboratory of Human Cytogenetics, University of Carthage, Tunis, Tunisia; 2) Department of Pediatrics, CHU Farhat Hached - Sousse-Tunisia; 3) Department of Genetics, CHU Farhat Hached - Sousse- Tunisia.

To date more than 50 cases of interstitial or terminal 8p23.1 deletions have been reported. This aberration is especially prone to various genomic rearrangements mainly because of the existence of the two oligofructose receptor gene clusters (REPD and REPP) which is associated with a spectrum of anomalies including especially congenital heart malformations microcephaly and mental retardation. We report on two Tunisian patients, with different clinical features but they have in common mental retardation, similar dysmorphic features, microcephaly otherwise without heart malformation. The conventional karyotype on R banded chromosomes at a resolution of 450-550 bands, revealed a terminal deletion of the short arm of chromosome 8 in only one patient. Using CGH array on a 75 Kb resolution average, we characterized this 8p deletion of 8,883,636 bp: arr 8p23.3- 8p23.1 (191,530- 9,075,165) X1. The second patient carried a smaller cryptic interstitial deletion of 1,181,530 pb at 8p23.1: arr 8p23.1 (8,229,404 - 9,410,934) X1. Fluorescence in situ hybridization (FISH) analysis confirmed the deletions and the parental analysis returned normal. Taken together, these data allowed us to define a critical deletion region of 845 Kb for the major features of a deletion 8p. Indeed, the absence of cardiac malformations in these cases, often associated with the 8p deletions, could be explained by the retention of SOX7 and GATA4 genes, implicated in heart development, in both parents. Also, we suggest that loci for the microcephaly and mental retardation are within the subregion of the commonly deleted region. Indeed, the different sizes of the reported deletions could otherwise explain the non-similarity of the clinical features observed in our patients. Even though, further studies are needed to establish more clearly the role of the discussed candidate genes in carriers of 8p23.1 deletions.
3205F Silver and Titanium dioxide Nanoparticles role in Genotoxicity. A. Gana-pathi1, L. Koudé1, V. Margarint1, P. Upendra2, Q. Hasan2, B. Priyanka1, R. Rohit1, R. Devaki1. 1) Biochemistry, Kamineni Institute of Medical Sciences, Hyderabad, Andhra Pradesh, India; 2) Department of Genetics, Kamineni Institute of Medical Sciences, Sreepuram, Nalgonda, Hyderabad, Andhra Pradesh, India; 3) Indian Institute of Chemical Technology, Hyderabad, Andhra Pradesh, India.

Objectives We investigated the genotoxic effects of 10-20 nm silver (Ag) and Titanium dioxide (TiO2) nanoparticles (NPs) by evaluating chromosome aberrations, polyplody status, Apoptosis and micronucleus (MN) assay. Methods Before testing, we confirmed that the Ag-NPs were completely dispersed in the ex-perimental medium by sonication (three times in 1 minute) and filtration (0.2 µm pore size filter), and then we measured their size in a zeta potential analyzer. After that the geno-toxicity was measured. Results After incubation with the defined concentration and size of the Silver and TiO2 nanoparticles, 100 metaphases were analyzed under microscope, 9%, 12% structural aberrations and 3%, 2% numerical aberrations were observed respectively. There was no aberration found in the control sample. After incubation with the defined concentration and size of the Ag and TiO2 nanoparticles, 100 metaphases were analyzed under microscope, 9%, 12% structural aberrations and 3%, 2% numerical aberrations were observed respectively. There was no aberration found in the control sample. However both nanoparticles did not show any apoptosis. Conclusions All of our findings indicate that Ag and TiO2 NPs show genotoxic effects in mammalian cell system. In addition, present study sug¬gests that the geno¬toxicity effect of Ag and TiO2 nanoparticles is concentration and size depend¬ent. Keywords Chromosomal aberrations, Polyploidy status, Genotoxicity, Toxicity effect of Ag and TiO2 nanoparticles is concentration and size depen¬dent.}

3206T Mosaicism involving a normal cell line and an unbalanced structural rearrangement. B.M. Shaw1, S. Eirefai1, B. Wolf1, L. Whiteley1, M. Strecker2, K. Havanes2, M. Cankovic1, M. Michalowski1. 1) Medical Genetics, Henry Ford Health System, Detroit, MI; 2) Pathology, Henry Ford Health System, Detroit, MI; 3) CombiMatrix Diagnostics, Irvine, CA.

Mosaicism for a normal cell line and a cell line with an unbalanced structural arrangement is a rare occurrence. We report a small-for-gestational age, female infant with cleft palate, low-set ears, ventriculomegaly and joint contractures born to a 28-year-old G2P2 mother. The parents are from Yemen and are reportedly non-consanguineous. The infant was intubated at birth for respiratory distress and eventually expired at 25 days of life due to respiratory complications. Chromosome analysis on cultured lymphocytes revealed a 46,XX,der(11)[1](1;1)[p36.3;q25.2][6]/46,XX[15] karyotype. Microarray analysis using a custom Illumina SNP array showed a female with a mosaic 1q25.3-q44 duplication and a 1p36.33-p36.31 deletion. These results were confirmed by both metaphase and interphase FISH analysis. These findings may be explained by one of four potential mechanisms: 1) chimer¬ism; 2) an unbalanced zygote, loses the abnormal chromosome 1 and undergoes monosomy rescue, resulting in isodisomy for the normal chromo¬some 1 in a subpopulation of cells; 3) a zygote with two normal chromosomes 1s and a derivate chromosome 1 is formed via 3:1 segregation from parent carrying a balanced inversion, and as a result of the abnormal cell line’s instability, secondary loss of the normal chromosome 1 and the derivate chromosome 1 occurs, resulting in two different cell lines; 4) a translocation occurs between chromatids during embryogenesis, resulting in either two unbalanced cell lines (one of which was non-viable), or a normal and a balanced cell line. Variable number tandem repeat (VNTR) analysis was performed and ruled out chimerism in this patient. Monosomy rescue is considered unlikely given that no stretches of homozygosity suggestive of isodisomy were detected by SNP array and the mother has a normal kary¬type. Although possible, A 3:1 segregation of chromosome 1 seems unlikely, since trisomy 1 has only once been reported in an eight-cell pre-embryo, and this configuration would also result in tetrasomy of a large portion of 1q. Based on these results, we speculate that the chromatid translocation during embryogenesis is the most likely mechanism.

3207F High resolution copy number analysis of genes involved in gonadal differentiation in patients with disorder of sexual development. S.A. Yatasekoro1,2, S. Nakajima3,4, F.X. Schnick1,2, J. Fox3, S. Madan-Khetarpal1,2,5, S.P. Wit¬cher1,2, A. Rajkovic1,2,7. 1) OB/GYN & Reproductive Sci, Univ Pittsburgh, Mage-Womens Hosp., Pittsburgh, PA; 2) Department of Pathol¬ogy, Univ Pittsburgh, School of Medicine, Pittsburgh, PA; 3) Department of Urology, Univ Pittsburgh; 4) Children’s Hospital of Pittsburgh of UPMC, PA; 5) Department of Pediatrics and Department of Medical Genetics; Univ Pittsburgh; 6) Division of Pediatric Endocrinology; Univ Pittsburgh; 7) Depart¬ment of Human Genetics, University of Pittsburgh School of Medicine, Pitts¬burgh, PA.

Cytogenetic and molecular genetic analyses are essential elements in clinical diagnosis and gender assignment in children with disorders of sexual development (DSD). In patients with multiple congenital anomalies the whole genome array comparative hybridization (aCGH) has become the first tier test in the clinical diagnosis. A significant number of genes have been recognized to play a role in male and female sex development, however small deletions and duplications involving an individual gene have remained beyond the detection resolution by routine aCGH. To detect copy number variations (CNVs) with the resolution of 0.5-1 kb we have constructed a custom CGH microarray with oligonucleotide probes covering 397 clinically relevant and candidate genes associated with gonadal differentiation and genitourinary tract development. We studied 37 patients diagnosed with sex-reversal or genitourinary tract abnormalities including seven XX males, seven XY females, and 23 individuals with 46,XY chromosome complement and partial testicular differentiation, ambiguous external genitalia, cryptorchi¬dism, or hypospadias. Pathogenic CNVs involving a single gene were detected in 4 patients (~10%) with nonsyndromic abnormalities. These include heterozygous and hemizygous imbalances within the CYP1B1, DAX1, POU3F, and FGF13 genes, ranging from 1 to 160 kb in size. The application of high-resolution custom DSD microarray in patients with DSD conditions provides substantial advantages in the detection of small genomic and intragenic aberrations that are beyond the resolution of currently avail¬able aCGH analysis and other molecular techniques such as FISH analysis and sequencing of the gene of interest. As the number of genes implicated in DSD development continues to grow, high resolution genome-wide approaches provide valuable information for the clinical diagnosis and man¬agement of patients with nonsyndromic DSD.

3208T 22q11.2 deletion in patients with syndromic cleft lip and palate. S. Carpenter1, I. Zarante2, M.C Martinez3, O.M Moreno1. 1) Biological science student Master. Institute of Genetics, Pontifical Javeriana University, Bogota, Colombia; 2) Teacher Institute of Genetics, Pontifical Javeriana University, Bogota, Colombia; 3) Teacher of Human Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA.

Craniofacial abnormalities are one of the most common features of all birth defects, orofacial disorders being the most frequent. Cleft lip and palate are the most prevalent craniofacial malformations. In order to detect 1 of every 700 newborns affected by this condition, we have to recognize the significant genetic and environmental factors and the role of genetic factors on variable deficiencies in tissues that form the palate and upper lip during the first weeks of gestation. In 30% of cases, genetic factors are found. Craniofacial abnormalities are described in 22q11.2 microdeletion syndrome as part of the 180 clinical signs known to contribute to the phenotypic spectrum. The aim of this study was to determine the presence of chromo¬somal abnormalities and 22q11.2 deletions in patients with cleft lip and palate and clinical suspicion of 22q11.2 deletion syndrome in patients from ECLAMC program and various hospitals in Bogota. The patients analyzed had cleft lip, cleft palate and cleft lip accompanied by other clinical features associated with 22q11.2 deletion syndrome, such as congenital heart dis¬ease, dysmorphic features, immunologic abnormalities or behavioral and learning disorders. We performed high-resolution G-banding karyotype and molecular analysis of the 22q11.2 region by MLPA (SALSA MLPA KIT P250-B1 DiGeorge). We evaluated 33 patients, 18 men and 15 women, 15 patients had palate clefts, 14 had cleft lip and palate and 3 had cleft lip. None of the patients had abnormal karyotype, 5 patients had the 22q11.2 deletion detected by MLPA, 4 of them had a deletion within the typical deleted region of ~3 Mb, and one had a atypical distal deletion expanding between LCR D and LCR F (~1.8 Mb). Patients with the confirmed deletion presented in addition to the cleft lip / palate, mainly cardiac abnormalities, immune abnormalities and abnormal facial features. The patient with the atypical deletion presented with hypothyroidism, anxiety, inguinal hernia and herpes zoster. These results suggest the importance of evaluating patients with cleft lip / palate in a genetics service and applying molecular tools as MLPA for analysis. The 22q11.2 region contains the genes involved in this group of abnormalities and provide appropriate genetic counseling.
3209F Co-occurrence of 22q11 Deletion Syndrome and HDR Syndrome. R. Fukai1,2, N. Ochi1, N. Matsumoto1, N. Miyake1. 1) Department of Human Genetics, Yokohama City Graduate School of Medicine, Yokohama, Kanagawa, Japan; 2) Department of Neurology and Stroke Medicine, Yokohama City University Graduate School of Medicine, Yokohama, Japan. We replicate our findings in a second large cohort of cases and controls. Using a custom CNV 'morbid map' was created using statistical/bioinformatic interpretations. The chromosomal regions of known CDH associated or CDH candidate contain CDH-causative genes. We analyzed aCGH data on 57 CDH patients however, the etiology remains unknown for the majority of cases. We have found chromosomal imbalances or point mutations have been implicated in CDH; occurring in isolation, and on occasion as part of a syndrome. Recurring events. Clinical, this study indicates the importance of carrying out careful clinical and genetic assessment of patients with atypical clinical phenotypes or unique combinations of symptoms and using whole genome copy number SNP arrays is especially useful for detecting such rare double mutations.

3210T Generation of a Custom Array for Copy Number Variation in Congenital Diaphragmatic Hernia. M. Russell1,2, K. Darvishi2,3, M. Longon1,2, CH. Hsieh2,3, F. High1, A. Kashani Pour1, A. Tracy1, C. Coletti1, C. Zhang2,3, B. Pober1,2, K. Lage1,2, C. Lee1,2, P. Donahoe1,2. 1) Dept Pediatric Surgery, Massachusetts General Hosp, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Department of Pathology, Brigham and Women's Hospital, Boston, MA. Congenital Diaphragmatic Hernia (CDH) is a common birth defect (1/3,000 live births in the United States), characterized by a diaphragm defect often with severe lung hypoplasia. CDH is genetically heterogeneous, usually occurring in isolation, and on occasion as part of a syndrome. Recurring chromosomal imbalances or point mutations have been implicated in CDH; however, the etiology remains unknown for the majority of cases. We have used array-based comparative genome hybridization (aCGH) to detect congenital diaphragmatic hernia (CDH) candidate loci, hypothesizing that they contain CDH-causative genes. We analyzed aCGH data on 57 CDH patients from our cohort with either isolated or complex CDH using two different platforms, Agilent 244k and Agilent 1M arrays. A total of 32 rare CNVs (defined as <5% frequency reported in the Database of Genomic Variants) were selected for follow up. We also combined our CNV findings with an additional 48 CDH-associated CNV regions reported in the literature, as well the chromosomal regions of known CDH associated or CDH candidate genes identified via different techniques in our laboratory. With these data a custom CNV 'morbid map' was created using statistical/bioinformatic integrations. This 'morbid map' of CDH-associated CNVs and genes can be validated by designing custom arrays on 350+ CDH patients within our cohort and 1,000+ controls, to identify CNVs that are significantly associated with CDH risk. In a collaborative effort with other CDH investigators, we will replicate our findings in a second large cohort of cases and controls. Using aCGH data, we have developed a morbid map and use a custom array to test statistically valid associations between rare CNVs and CDH.

3211F Multiplex Ligation-dependent Probe Amplification (MLPA) for detection of copy number variation in early-onset glaucoma patients from the US, K. Allen, M. Janssens, K. Lkhurkhuu, W. Abdurabou, E. DeiBono, J. Wiggins, Massachusetts Eye and Ear Infirmary, MA. Early-onset glaucoma can be caused by mutations in CYP1B1, FOXC1, and PITX2. We have previously identified mutations in these genes in approximately 1/3 of patients with early-onset glaucoma. Using MLPA, we have performed MLPA testing on genomic DNA samples from 50 early onset glaucoma patients using the MRC Holland SALSA MLPA probemix kits for CYP1B1 (P128-B2 Cyclotide P-450) and FOXC1 and PITX2 (P054-B1 FOXL2-TWIST1). MLPA raw data analysis was analyzed using the MRC Holland protocol for internal quality controls. Intra-sample normalization was calculated by comparing the relative florescence intensity of each peak within a sample. Inter-sample normalization was calculated by dividing the relative peak area of each experimental sample to that of reference samples included within the same experiment. In this study, a normal dosage quotient was 0.85-1.15. A dosage quotient of 0.35-0.65 was considered a heterozygous deletion. MLPA analysis for this study revealed heterozygous deletions in CYP1B1 (1 patient), FOXC1 (3 patients) and PITX2 (1 patient). The heterozygous CYP1B1 deletion found in this study spanned the entire coding region for this gene. Insertions or homozygous copy number variants were not identified. The results of this study show that deletions in CYP1B1, FOXC1, and PITX2 are relatively frequent mutations in this patient population and should be included as part of the genetic diagnostic evaluation.

3212T Autoimmune Disorders as Late Onset Feature of 16p11.2 Duplications. C. Lowther1,2, D.J. Stavroupolous2,3, S. Dyack4,5, AL. Lionel6, CR. Marshall7, SW. Scherer1,2, AS. Bassett1,2, 1) Institute of Medical Sciences, University of Toronto, Toronto, Ontario, Canada; 2) Cytogenetics Laboratory, Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, Canada; 4) Department of Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada; 5) Izaak Walton Killam Centre, Division of Medical Genetics, Halifax, Nova Scotia, Canada; 6) The Centre for Applied Genomics and Program in Genomics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 8) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. The use of genome-wide microarrays to detect pathogenic copy number variations (CNVs) is now routine in many postnatal settings, and there is growing interest in the application to prenatal diagnosis. This has placed additional pressures on the medical genetics community to delineate the variable expressivity of CNVs such as recurrent 600 kb 16p11.2 duplications to inform genetic counselling and anticipatory care. Notably, there is limited information as yet on the adult phenotype apart from the known enrichment of 16p11.2 duplications in schizophrenia (SZ). We assessed multiple relatives of probands previously identified to carry the 16p11.2 duplication from two cohorts comprised of adults diagnosed with SZ or tetralogy of Fallot. Each individual, regardless of duplication status, underwent phenotyping of major and minor neuropsychiatric features. We report on the first 19 adults from five families, 12 of whom had the 16p11.2 duplication (median age 58, range 32-88). Eight (66.6%) of these 12 had some form of neuropsychiatric illness. In the 7 patients who carried the duplication, including two individuals with severe depression and one with intellectual disability, in addition to the four ascertained with SZ. None of the seven individuals had spinal abnormalities: spina bifida occulta and spinal stenosis. There were also five with autoimmune disorders including rare conditions: lichen planus, myasthenia gravis, and Sjögren's syndrome. One individual had developed severe nephrolithiasis, a rare form of dystonia, in middle age. To our knowledge this is the first study to report that late onset autoimmune and movement disorders may be features of the extended adult 16p11.2 duplication phenotype. With respect to the clinical phenotype of the 16p11.2 duplication, two individuals had spinal abnormalities: spina bifida occulta and spinal stenosis. There were also five with autoimmune disorders including rare conditions: lichen planus, myasthenia gravis, and Sjögren's syndrome. One individual had developed severe nephrolithiasis, a rare form of dystonia, in middle age. To our knowledge this is the first study to report that late onset autoimmune and movement disorders may be features of the extended adult 16p11.2 duplication phenotype. With respect to the clinical phenotype of the 16p11.2 duplication, two individuals had spinal abnormalities: spina bifida occulta and spinal stenosis. There were also five with autoimmune disorders including rare conditions: lichen planus, myasthenia gravis, and Sjögren's syndrome. One individual had developed severe nephrolithiasis, a rare form of dystonia, in middle age.
Evaluation of newborns with multiple congenital anomalies by array-CGH: the experience of a public hospital from Southern Brazil. M. Roeger1,2,3, L. Dorfman1, R. Mergener2, K. de Souza3, J.C. Leite3. 1) Post Graduate Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil; 2) Molecular Cytogenetics Laboratory, Gene Therapy Center, Hospital de Clínicas de Porto Alegre, Porto Alegre-RS, Brazil; 3) Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre-RS, Brazil.

Aim and methods: We analyzed DNA samples isolated from individuals with multiple congenital abnormalities born in a public hospital from Southern Brazil to identify copy number variants involving critical chromosome regions using array comparative genomic hybridization (array-CGH). Results: In a series of samples from 32 newborns with Multiple Congenital Anomalies (MCA) we identified six deletions and four duplications. We identified one subject with a previously unknown 12 Mb deletion at 4q16.3-p15.33 associated with the Wolf-Hirschhorn syndrome region and one subject with a complete trisomy of chromosome 13. Thirty of the syndromic cases had MCA without a specific clinical diagnosis. The deletions were pathogenic and associated with the phenotypes in two cases and the duplications were classified as pathogenic and associated with the phenotypes of two subjects. In two further cases, an observed genomic imbalance was classified as variation of unknown significance (VOUS) as there was insufficient evidence to conclude that the CNV was either pathogenic or benign. In general, gain involving CNVs was more common than loss, accounting for ~70% of the imbalances. Conclusion: We have shown that array-CGH analysis of DNA samples obtained from newborns with MCA of unknown etiology is an efficient method to identify candidate chromosomal loci and/or genes, complementing the genetic evaluation of newborns with MCA. Cytogenetic testing is expected to use mostly array-CGH based technology in the near future. However, the high cost associated with the cytogenetic molecular methods and the present lack of technical skills and professional experience needed for its application are major challenges for public hospitals in developing countries such as Brazil. Financial support: FIP/HCPA 10-560; CNPq 214906/2012-4; CNPq 402012/2010-0.

3214T
Germline DNA Copy Number Variation investigation in individuals with Argyrophilic Grain Disease reveals CTNS as a plausible candidate gene. C. Rosenberg1, L. Kimura1, D. Schlesinger2, A. Gonçalves3, P.L. Pearson1, C.K. Suemoto4, C. Pasquaulli5, A.C. Krepischi1, L.T. Grinberg1, D. Vieira1. 1) Genetics Evolutionary Biology, Universidade de Sao Paulo, sao paulo, sao paulo, brazil; 2) Israel Institute for Teaching and Research Albert Einstein., Sao Paulo, Brazil; 3) National Institute of Science and Technology in Oncogenomics, AC Carmargo Hospital, Sao Paulo, Brazil; 4) Discipline of Geriatrics, Department of Internal Medicine, University of Sao Paulo Medical School,Sao Paulo, Brazil; 5) Department of Pathology, University of Sao Paulo Medical School,Sao Paulo, Brazil; 6) Brazilian Aging Brain Study Group - LIM22, Department of Pathology, University of Sao Paulo Medical School, Sao Paulo, Brazil.

Argyrophilic grain disease (AgD) is a late-onset dementia morphologically characterized by the presence of argyrophilic grains (AgR) in neuronal processes and coiled bodies in oligodendrocytes. Several studies have confirmed that the main protein constituent of AgR is the abnormally hyperphosphorylated tau protein and that the grains are found in both cortical and subcortical structures. It also became apparent from recent clinico-pathological studies that AgD shows a significant correlation with advancing age and it may be present for approximately 5% of all dementia cases. However, the cause of AgD is not known. The disease seems to be sporadic but genetic studies have failed to discover a link of AgD with a particular gene locus. An important and recent advance in human genetics was the recognition that our genome presents copy number variations (CNVs) that involve gain or losses of genomic DNA among phenotypically normal individuals. Indeed, it has become well established in the literature that CNVs represents a significant proportion of the total genetic variability in the human population. Since the discovery of the existence of CNVs it has increased the number of studies that demonstrate its critical role in the phenotypes of complex diseases. However, there are no reports of specific CNVs related to AgD. Therefore, the aim of this work was to identify CNVs, using the array-CGH, that possibly contributes to the development of AgD. The samples were obtained from the Brain Bank of the Brazilian Aging Brain Study Group. The array-CGH analysis revealed a 315 Kb genomic imbalance in the 17p13.2 region that includes a gene involved in age-related memory deficit, the CTNS. There is a study showing that a knockout mice Ctns --/- presents a severe age-related memory deficit, which makes CTNS a good candidate gene for susceptibility to AgD.
Delineation of the genomic structure in the human 2q13 region. B. Yuan1, P. Liu1, K. Poliamousis2, D. Schwartz2, J.R. Luski1,1,1, 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Laboratory for Molecular and Computational Genomics, University of Wisconsin-Madison, Madison, WI 53706, USA; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 4) Texas Children's Hospital, University of Texas Health Science Center at Houston, Houston, TX 77030-1501, USA.

Familial juvenile nephronophthisis is an autosomal recessive kidney disorder, which is the most frequent inherited cause of chronic renal failure in children. Loss of function of the gene NPHP1 is responsible for approximately 85% of the nephronophthisis cases. About 80% of these patients carry a large homozygous deletion involving NPHP1, NPHP1 maps to 2q13, a region with extremely enriched low copy repeats (LCRs). Previous studies have demonstrated that deletion of the NPHP1 can be mediated by nonallelic homologous recombination (NAHR) between two 45kb flanking repeats. One non-pathologic inversion haplotype involving NPHP1 region flanked by two 385kb large inverted repeats has also been reported. Moreover, extensive evidence suggest that the human 2q13 region is structurally highly polymorphic. Therefore, we hypothesize that different historical genomic rearrangements occurred within the 2q13 region and generated various structural haplotypes observed in the human population today. Individual susceptibility to the NPHP1 deletion can be affected by the haplotypes they carry in their personal genomes. By exploring the different haplotypes from various individuals, we may glean a more detailed understanding of the correlation between disease susceptibility and structural variation mediated by LCRs. We aim to find evidence to support the various predicted structural haplotypes. By mining the data from the literature, we observe the polymorphic copy number variation of the 45kb repeats in the 2q13 region in different populations. Optical Mapping analysis of one HapMap individual reveals the inversion haplotype of NPHP1 region as well as the deletion of one 45kb region. We have observed 25 discordant fosmid clones constructed in the Human Genome Structural Variation (HGSV) project to infer potential haplotypes. The discordant fosmid clones whose mapping cannot be explained by the 2q13 genomic structure in the reference genome imply alternatively structured chromosomes.

Designing Custom Oligo FISH Probes for the Detection of Chromosomal Rearrangements in FFPE Tissues. A. Bergstrom Lucas1, M. Ruvolo1, V. Kulkarni1, S. Chen1, B. Mullinax2, J. Vennen2, J. Barboza1, S. Happe2, S. Fulmer-Smentek1, M. Srinivasan1. 1) Genomics R&D, Agilent Technologies, Santa Clara, CA; 2) Genomics R&D, Agilent Technologies, La Jolla, CA; 3) Genomics R&D, Agilent Technologies, Cedar Creek, TX. Cancer cells frequently contain chromosomal rearrangements that result from oncogene activation, and the genes involved in these rearrangements are increasingly being identified using molecular technologies. We have developed a new generation of fluorescently labeled in situ hybridization (SureFISH) probes for the detection of these rearrangements. SureFISH probes are comprised of thousands of unique long oligonucleotides that are tiled across the targeted chromosomal region avoiding non-unique portions of the genome. The oligonucleotides are synthesized using Agilent’s Oligonucleotide Library Synthesis (OLS) technology. Using knowledge of translocation breakpoints, SureFISH probes are designed to detect the translocated sequences using both break-apart and dual fusion strategies. The in silico design methodology and de novo synthesis of the SureFISH probes enable the optimization of design characteristics so that each probe provides balanced signals, facilitating the detection of chromosomal rearrangements. The flexibility afforded by the SureFISH design pipeline also enables rapid probe customization. Custom designs can be generated that target almost any genomic region, allowing for the production of probes that are not possible using other methods. We demonstrate the performance of both catalog/routine and custom probes on cytological samples and tissues that have been preserved in formalin and embedded in paraffin (FFPE).
**3220T**


The use of pesticides is still the main strategy to fight plagues in agriculture. Methamidophos is an organophosphorus pesticide largely used in the world crop due to its efficiency. The purpose of the present work was to study the effect of methamidophos on chromosomes of mice bone marrow cells in vivo, using the micronucleus assay. Animals (ICR mice) were separated into four groups. In the first group, 6 animals received methamidophos intraperitoneally during five consecutive days in a concentration equivalent to 25% of the LD50. In the 6th day, animals were sacrificed, their femurs removed, the bone marrows collected and smears were made for slides preparation. After 24h cells were stained with Giemsa Gurr (2%) and analyzed under optical microscope. As positive control, 6 animals received cyclophosphamide (50mg/mL), once and 6 animals not exposed to any drug served as negative control for the experiment. In the test group, 13353 cells were observed and 183 showed micronuclei. In the positive control group, 12037 cells were observed and 238 had micronuclei and in the negative control group, of 12156 cells observed, none had micronuclei. The chi-square test for independence showed that our results were extremely significant (P<0.0001). They suggest that methamidophos is responsible for the micronuclei observed.

**3221F**

Mood disorders associate with a gene-disrupting chromosomal translocation spanning three-generations. C. Ernst1,2, F. Jollant1, C. Hanscom2, M. Stone3, I. Blumenthal2, G. Turecki1,2, C. Cruceanu1, M. Tałkowski3,4, 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Department of Psychiatry, McGill University, Montreal, Quebec, Canada; 3) Center for Human Genetic Research, Harvard Medical School, Boston, MA, USA; 4) Medical and Population Genetics program, the Broad Institute, Cambridge, MA, USA.

We identified a family that carried a translocation across three generations [46,XX,t(11;14)(p11.2;p12)]. Due to a profound history of mood and anxiety disorders associated with the anomaly, we sequenced the translocation to basepair resolution using a mate-pair library preparation followed by massively parallel sequencing. The translocation breakpoints were localized to an unmapped contig, chrUn_gi000220, which we can now locate to Chr 14p12, and to intron 3 of LRRC4C. LRRC4C is a binding partner for Nethrin G1, an axon guidance molecule important in neurodevelopment. To understand if LRRC4C is important in mood disorders, we analyzed exome sequencing data from a large bipolar family study and identified additional subjects with very rare amino-acid altering mutations in this gene, suggesting that LRRC4C mutations might have a role depressed mood and anxiety. These data demonstrate the benefits of using whole-genome sequencing to detect structural variation for psychiatric disorders and for localizing unmapped contigs to the human genome.

**3222T**

Effectiveness of comprehensive cytogenetic investigations in the diagnosis and genetic counseling of 6,131 patients. L. Martelli1,2, J. Huber1, S.A. Santos1, L.A.F. Laureano1, A.C. Laus1, J.A. Squire1, E.S. Ramos1,2, 1) Dept Genetics, University of Sao Paulo, Ribeirao Preto, SP, Brazil; 2) Clinical Hospital, School of Medicine, Ribeirao Preto, SP, Brazil; 3) Dept Pathology and Molecular Medicine, Queen’s University, Kingston, ON, Canada.

Aberations involving gain or loss of chromosomal segments are particularly important in clinical cytogenetics. Classical karyotyping systematically supported by companion molecular cytogenetics has been a powerful tool for genome scanning, detecting aberrations involving gains and losses of chromosomal regions, as well as intra and inter-chromosomal rearrangements. The main objective of this study was to demonstrate the importance of comprehensive chromosomal and molecular cytogenetic analysis for clinical diagnosis over the last 15 years at the Department of Medical Genetics (HCFMRP). Initially, we evaluated 8,131 karyotypes from patients, of these 7,469 were informative, corresponding to 91.85% of the samples. Chromosomal aneuploidies were detected in 1,363 karyotypes (18.25%), structural aberrations in 405 karyotypes (5.42%) and 483 karyotypes (6.47%) had normal karyotypes with variants of normality. Among the translocations, 26 were inherited and these findings were used to determine the risk of recurrence. A subset of 17 patients required extensive molecular cytogenetic analysis including SKY and FISH techniques, to evaluate supernumerary marker chromosomes. The definitive diagnosis was established for 71% of these patients. Seven markers were originated from acrocentric chromosomes - five derivatives of chromosome 15 and two from 22-, two from chromosome 8, and one from each pair 2, 4, 10 and 18. For four cases the origin was inherited. For two cases mBand analysis provided additional characterization of the chromosomal rearrangements. In one case of reciprocal translocation (1q;11p), small deletions were excluded in 1q31 and 11p13 cytobands. In the second case, the karyotype with homozygous inversion in chromosome 12 was characterized as 46, XY, der (12) inv (12) (p11.2p12.3) inv (12) (q21.1q24.1) x2. This patient was also investigated by aCGH technique which showed genomic gains within cytoband 12p12.3 that seemed to be related to the clinical phenotype. In conclusion, our results suggest that G2 phase cells with sub-threshold DNA damage enter mitosis due to failure of the ATM-dependent G2 arrest. We hypothesized that the cytogenetic dose-response relationship at very low doses is also non-linear, and hence, should not be extrapolated from the effect at high doses. To test our hypothesis, experiments were performed in which human peripheral blood lymphocytes from one normal healthy blood donor were acutely exposed to cobalt-60 gamma rays at doses 0-4 Gy, with dose groups below 1 Gy spaced at intervals of 0.1 Gy. Damage was assessed using structural chromosomal aberrations. Preliminary results indicate that the effect per unit dose is up to 6-fold higher between 0 and 0.3 Gy than between 0.4 and 1.5 Gy. These data provide the first cytogenetic evidence for the existence of LDH in human peripheral blood lymphocytes and suggest that the LNT model is not optimal for making radiation risk assessments in the low-dose region.

**3223F**

Characterizing the cytogenetic dose-response relationship at very low doses of gamma radiation using structural chromosomal aberrations. L. Seth1, M.C. Joiner2, J.D. Tucker1, 1) Biological Sciences, Wayne State University, Detroit, MI; 2) Department of Radiation Oncology, Wayne State University, Detroit, MI.

The shape of the dose-response curve at very low doses of radiation has long been a topic of debate in the field of radiation biology. The linear-no-threshold (LNT) model is widely used to assess risks associated with low dose exposures. According to the LNT model, the risks of genetic damage increase linearly with dose without any threshold. However, lack of knowledge of the shape of the dose-response curve at low doses, where most exposures occur, is a major problem because if low dose risks are not understood, then risk estimation and establishing scientifically valid dose limits for radiation protection are not possible. The low-dose hypersensitivity (LDH) phenomenon, in which cells are especially sensitive to low doses but then show increased radio-resistance at higher doses, provides evidence of non-linearity in the low dose region. LDH is more prominent in the G2 phase of the cell cycle than the G0/G1 or S phases and others have shown that G2 phase cells with sub-threshold DNA damage enter mitosis due to failure of the ATM-dependent G2 arrest. We hypothesized that the cytogenetic dose-response relationship at very low doses is also non-linear, and hence, should not be extrapolated from the effect at high doses. To test our hypothesis, experiments were performed in which human peripheral blood lymphocytes from one normal healthy blood donor were acutely exposed to cobalt-60 gamma rays at doses 0-4 Gy, with dose groups below 1 Gy spaced at intervals of 0.1 Gy. Damage was assessed using structural chromosomal aberrations. Preliminary results indicate that the effect per unit dose is up to 6-fold higher between 0 and 0.3 Gy than between 0.4 and 1.5 Gy. These data provide the first cytogenetic evidence for the existence of LDH in human peripheral blood lymphocytes and suggest that the LNT model is not optimal for making radiation risk assessments in the low-dose region.
3224T
Double Robertsonian translocation in pregnancy loss. S. Wenger Dept Pathology, West Virginia Univ, Morgantown, WV.

The incidence of balanced Robertsonian translocation is 1/1,000 with rob(13;14) being the most common at 76%, followed by rob(14;21) at 10%. Data from our laboratory during the past 26 years identified 64 cases, two of which were acquired in leukemia and one which was mosaic with a normal karyotype. The most common Robertsonian translocation was 13:14 (52%) followed by 14:21 (16%). Absent from our records were 14:22, 21:22 and 22:22. Fourteen families were identified in which eleven demonstrated maternal inheritance, two were paternal and one case had a noted family history of trisomy 14 with no further detail. It was not possible to determine the de novo rate since not all parents were evaluated for all of the Robertsonian translocations. Adverse outcomes involving trisomies related to the Robertsonian translocation included 13, 14, and 21, although one case had an additional three way reciprocal translocation and another case had a single X chromosome. Within these fourteen families, not all pregnancy losses were karyotyped, however one pregnancy loss had both trisomy 13 and 15 as the result of two copies of rob(13;15)mat. Only one other case has been reported in the literature with a duplicated Robertsonian translocation; a karyotype of 44.XX,rob(14;21)(q10;q10)mat2 in a 17 year old woman with primary amenorrhea (Kopakka et al., Int J Gynecol Obstet 2012;116:253-257). The most likely explanation for two copies of an inherited Robertsonian translocation is nondisjunction of the Robertsonian translocation in meiosis II or first mitotic division of the zygote.

3225F
Detection of Interstitial 9q34.2-q34.3 Deletion Including the COL5A1 Gene by Array-CGH in a Patient With Classical Ehlers-Danlos Syndrome and Developmental Delay. E.C. Beltran1,2,3, D.W. Stockton1,2,3, M. Hicks1,2,3, M. Hankerd1,2,3, D. Schloff1,2,3, M. Kristofice4,5, S. Ebrahimi1,2,3, 1) Children’s Hospital of Michigan-Specialty Center, Detroit, MI; 2) Division of Genetic and Metabolic Disorders, Detroit, MI; 3) Department of Pediatrics, Wayne State University School of Medicine, Detroit, MI; 4) Department of Pathology, Wayne State University School of Medicine, Detroit, MI; 5) Detroit Medical Center University Laboratories, Cytogenetics Laboratory, Detroit, MI.

Ehlers-Danlos syndromes (EDS) are a heterogeneous group of heritable connective tissue disorders. The classic type is characterized by hyperextensibility, skin atrophy, cutaneous scars due to tissue fragility and joint hyperlaxity. Mutations within the COL5A1 (9q34) or COL5A2 (2q31) genes have been found in about half of patients with Classical Type EDS. We report a 13 year old female who was referred for a genetic consultation because of EDS, developmental delay with mild cognitive impairment and partial complex seizures. She was diagnosed with EDS at 8 years by a skin biopsy. Genetic workup for a unifying diagnosis identified an interstitial 9q34.2-q34.3 deletion encompassing the COL5A1 gene. Array comparative genomic hybridization (CGH) testing detected a 1.95 Mb interstitial deletion involving chromosome 9 long arm region: arr[het]18 9q34.2q34.3 (136,405,205-138,358,850)x1. The deleted segment harbors 35 genes including COL5A1 gene implicated in Ehlers-Danlos syndrome. However, the deleted region does not include the EHM1 gene suggested to be associated with 9q subtelomeric deletion also known as Kleefstra syndrome. Partial deletions of chromosome 9 long arm are rare, and are associated with a wide range of phenotypes depending on the size of the deleted region. A MEDLINE search for reports of chromosome 9q34.2-q34.3 deletions encompassing the COL5A1 gene did not reveal similar cases. However, there are a few reported cases with deletions that overlap this region which are described in individuals with variable degrees of developmental delay, cognitive impairments, facial dysmorphism, and other congenital malformations. Deletion of chromosome 9q encompassing the COL5A1 gene in our patient is likely satisfactory explanation for her classic EDS presentation plus her neurological manifestations. The specific contribution of the other genes in the region to her neurologic symptoms is not clear. She has some phenotypic similarities to the 9q subtelomeric deletion syndrome despite excluding the EHM1 gene.

3226T
Use of cytogenetic tools to detect oxidative damage markers in diabetes mellitus and cardiomyopathy patients. R. Saraswathy1, C.R. Mathew1, K.T. Babu2, 1) Biomolecules and Genetics, VIT University, Vellore, India; 2) Madras Medical College, Chennai; 3) Heartline Medical and Research centre, Vellore.

Oxidative DNA damage is known to be strongly correlated with age related diseases like diabetes mellitus and cardiomyopathies. Various in vitro and in vivo cytogenetic methods have been developed to assess the DNA damages in human and animals. In this study a correlation between the DNA damages in diabetes mellitus and cardiomyopathy patients due to oxidative stress developed during pathogenesis was determined. The cytogenetic tools such as chromosomal aberration analysis and cytokinesis block micronucleus cytome assays were utilized. The results revealed a significantly higher frequency of DNA damage in the patients than in the controls. The interaction between the defect in mitochondrial DNA (mtDNA) and chromosome anomalies and other factors that they are exposed to may lead to DNA damage and genomic instability. This study will not only lead to the use of cytogenetic tools to assess the oxidative damage markers in other human diseases but develop interventional strategies for prevention.

3227F
Partial Trisomy 21 in a Patient without the Down Syndrome Phenotype. A. Umrigar1, T.J. Chen2, F. Ties1. 1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA; 2) Hayward Genetics Center, Tulane School of Medicine, New Orleans, LA.

We present a 24-year old female who was found to have trisomy 21 by routine karyotype, but did not present with the typical Down Syndrome phenotype. The patient had short stature, strabismus, and mental deficiency, but none of the common dysmorphic features of Down syndrome. Previous results indicated that the patient had three copies of chromosome 21, one of which appeared smaller in size. Subsequent fluorescence in situ hybridization (FISH) using the Down Syndrome Critical Region (DSCR) probe showed a deletion of the 21q22.11-22.2 region on the third copy of the 21 chromosomes. Microarray comparative genomic hybridization (aCGH) revealed a deletion of approximately 13.72 Mb at 21q22. Thus, it is concluded that the patient’s lack of the Down Syndrome phenotype is due to this deletion. This occurrence is extremely rare, requiring both the deletion of the DSCR and a meiotic non-disjunction event of the normal chromosome. Additionally, the proband later miscarried a fetus with Potter syndrome, but we were unable to perform cytogenetic and molecular analyses of the products of conception (POC). Our case demonstrates the importance of molecular confirmation in atypical patients previously diagnosed by karyotype.

3228T
Age-related decrease of meiotic cohesins in human oocytes. M. Tsutsui1, R. Fujiiwara2, H. Nishizawa3, H. Kogo4, H. Inagaki5, T. Ohye5, T. Kato1, H. Kurahashi1. 1) Fujita Health University, Toyoake, Japan; 2) Gunma University, Maebashi, Japan.

Chromosomal segregation error, e.g. nondisjunction or predivision, during meiotic cell division produces aneuploid gametes. Aneuploidy of fetal chromosomes causes miscarriage or newborns with congenital birth defects such as Down syndrome (DS). It is well known that the majority of the extra chromosomes of trisomy 21 in DS originates from segregation error of maternal meiosis I, and the other one half from paternal meiosis I. Age-related aneuploidy causes elevated risk of miscarriage in older women. However, the etiology of maternal age-related increase of segregation error remains unclear.

Meiotic cohesins have a key role in correct segregation of chromosomes in meiosis. In mammalian oocytes, the cohesion between sister chromatids is established by the cohesin complex at fetal stage, and then the oocytes become arrested at prophase I for a prolonged period until they are ovulated after the sexual maturation. However, it has been shown that the meiotic cohesin does not undergo turnover after birth in female mice. It is possible that this might cause the gradual decrease of cohesins, and then results in increased risk of age-related segregation error.

To verify this hypothesis, we examined the amount of meiotic cohesins in oocytes by quantification of immunofluorescence signal intensity on ovarian tissue sections. Oocytes were obtained from ovaries surgically resected from eight women having ovarian cancers (age range: 19-49 years) as well as those from normal C57BL/6N female mice. It was demonstrated that the immunofluorescence levels of meiosis-specific cohesin subunits, RECA and SMCP1, decreased in women in their 40s compared to around 20 years old. Age-related decrease of meiotic cohesins was also shown in mice. In contrast, the signal levels of subunits shared by mitotic and meiotic cells, SMCP3 and RAD21, did not change. These results suggest that the decrease of meiotic cohesin subunits impairs the cohesion between sister chromatids with age to induce segregation error. To elucidate the mechanism of decrease of meiotic cohesins will contribute to develop a novel strategy to prevent age-related increase of aneuploidy.
Microcephaly and chromosomal abnormalities: review of four years study. M. Kammoun1, S. Dimassi1, H. Hannachi1, I. Bel hadj Hmidia1, N. Soyah1, H. El Ghziel1, A. Saad1, M. Soumaya1. 1) department of cytogenetic and reproductive biology, Farhat hached, university teaching Hospital, Sousse, Tunisia; 2) Department of pediatrics, Farhat hached, university teaching Hospital, Sousse, Tunisia.

Microcephaly is clinically a heterogeneous disorder caused by several conditions including genetic defects. We analyzed cytogenetic profiles of 135 children referred for microcephaly during the period from January 2009 to December 2012. Standard karyotyping was abnormal in 14% of cases. Conventional and molecular Cytogenetic analysis using Fluorescent In Situ Hybridization and Comparative Genomic Hybridization array showed us that microcephaly is mostly related to Angelman syndrome (42%), chromosome 18 rearrangements (26%) and Wolf Hirschon syndrome (WHS) secondary to 4p16.3 deletion (21%). Down syndrome, Williams and Turner syndrome, 8p23.1 and 5p21 deletion, 17q23 duplication, trisomy X and Xq28 duplication were also associated to microcephaly. Microcephaly is commonly induced by autosomal recessive mutations. Nevertheless, it is associated to the majority of chromosomal rearrangements. UBE2A gene, which is involved in AS, is proved to contribute to cellular proliferation and its invalidation was shown to be significantly associated to increased cerebellar apoptosis. Microcephaly is probably related to DOK6 haplinsufficiency located at 18q 22 and shown to be involved in cortex neurite outgrowth and in neural precursor cell differentiation during cortical development. Otherwise, Xq28 duplication is an emerging syndrome that involves MECP2 gene, a transcriptional regulatory factor strongly involved in the central nervous system development. Microcephaly is frequently associated to chromosomal rearrangement. CGH array is of great importance to detect cryptic anomalies and to narrow critical regions of microcephaly and so identifying new candidate microcephaly genes.

Constitutional 560.49 kb chromosome 2p24.3 duplication including the MYCN gene identified by oligonucleotide and SNP chromosome microarray analysis in a child with multiple congenital anomalies and bilateral Wilms tumor. M.A. Micale1-3, B. Embrey IV1, J.K. Mackinis1, C.E. Harper2, D.J. Aughton1. 1) Dept Pathology and Laboratory Medicine, Beaumont Health System, Royal Oak, MI; 2) Beaumont Children’s Hospital, Royal Oak, MI; 3) Oakland University William Beaumont School of Medicine, Rochester, MI.

Less than 100 patients with partial chromosome 2p trisomy have been reported. Clinical features are variable and depend on the size of the duplicated segment, but generally include psychomotor delay, facial anomalies, congenital heart defect, and other abnormalities. A subset of patients has also developed neuroblastoma. We report a 560.49 kb duplication of chromosome 2p in a 13 month-old male with hydrocephaly, ventricular septal defect, partial agenesis of the corpus callosum, and bilateral Wilms tumor. He underwent neoadjuvant chemotherapy followed by right radical nephrectomy that revealed triphasic Wilms tumor with prominent histological and skeletal muscle differentiation. A needle core biopsy on one of two lesions on the left kidney revealed Wilms tumor. A partial left nephrectomy revealed focally positive margins that necessitated left flank radiotherapy. The tumor genotype was 46,XY,i(7)(q11.2),+2p,6,XY[12]. DNA sequencing of the WT1 gene revealed no mutation. His constitutional karyotype was 46,XY, suggesting that the i(7)(q36;p11) was associated with the malignancy. Cytogenomic evaluation of peripheral blood utilizing a 135K Nimblegen oligonucleotide microarray with conventional copy number duplication involving four OMIM genes: NBAS, DDX1, MYCNOS, and MYCN. SNP array analysis of the tumor utilizing the CytoScan HD array identified a 560.49 kb chromosome 2p24.3 duplication with no other pathogenic genomic abnormalities. This is, to our knowledge, the first partial trisomy 2p case with constitutional MYCN duplication in a child with bilateral Wilms tumor. As of May 2013, the 2-year-old boy continues to do well without clinical or radiographic evidence of recurrent disease. This case is also instructive because the child’s health insurer initially denied authorization for CMA on the basis that there is insufficient evidence in the peer-reviewed literature to demonstrate the clinical/therapeutic utility of CMA, and it was more than one year before such authorization was granted. That initial decision to deny coverage could have had untoward health implications for this child, as the identification of constitutional MYCN duplication necessitated surveillance for neuroblastoma. While the child did not develop neuroblastoma during the time it took to get insurance authorization for CMA, it is obvious how such a delay could have precluded early detection of a second malignancy for which this child is at high risk.

Incorporation of flapon probes reduces truncation losses for fluorescence in situ hybridization analysis of recurrent genomic deletions in tumor sections. M. Yoshimoto1,2, O. Ludkovski1, J. Good1, R.J. Gooding2, A. Boag1, A. Evans2, M.S. Tsao3, P. Nunn1, J. McGowan-Jordan1,2, J.A. Squire1. 1) Department of Pathology and Molecular Medicine, Queen’s University, Kingston, Canada; 2) Cytogenetics Laboratory, Children’s Hospital of Eastern Ontario, ON, Canada; 3) University Health Network, Princess Margaret Hospital, Division of Applied Oncology, Toronto, ON, Canada; 4) Department of Physics, Engineering Physics and Astronomy, Queen’s University, Kingston, ON, Canada; 5) Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada.

The establishment of quality control standards in clinical laboratories using routine fluorescence in situ hybridization (FISH) analysis of formalin-fixed paraffin-embedded (FFPE) sections is crucial as more genomic biomarkers are used for diagnostic and prognostic evaluation. For gene fusion, break-apart and gene amplifications there are guidelines for both analysis and reporting, but at the present time there are no comparable guidelines for the application and analysis of genomic deletions using FPE sections. The use of FISH on archival FFPE samples is technically demanding and becomes more challenging when applied to paraffin-embedded tissue microarrays. In this study we report a generalizable four-color deletion FISH approach to assist interpretation problems arising when evaluating FISH signals. The guidelines will help address interpretative dilemmas associated with overlapping and truncated nuclei in FFPE prostate cancer sections. The four-color FISH approach was developed using the PTEN gene deletion model in prostate cancer. The PTEN assay was based on a robust bioinformatics analysis of 311 published human genome array datasets and comparative analysis of the four-color assay with a specific PTEN probe, and flanking probes either side of the target probe. The sensitivity and specificity parameters of the four-color PTEN probe set were further characterized using a large number of well-characterized tumors and stringent scoring criteria. The incorporation of flanking probes in the analyses to determine if the chromosomal region was subject to truncation loss. A minimum threshold for apparent deletion frequency was set to address the heterogeneous and homogeneous nature of tumor histology. In addition the approach facilitated analysis of genomic heterogeneity and varying clonality within different foci of tumor in the prostate. Overall the approach provided robust and highly reproducible results that minimized inter- and intra-assay variability. The four-color FISH deletion assay reduced the frequency of misinterpretation and misestimation. For gene fusion, break-apart and gene amplifications there are guidelines for both analysis and reporting, but at the present time there are no comparable guidelines for clinical importance are discovered by next generation sequencing methods.

High resolution genome profiling in Li-Fraumeni patients without germ-line (LFS/Basso et al., 2016) and without (LFS) syndromes predispenses a variety of different tumors occurring over a wide age range and are associated with germline mutations in TP53 gene. Peripheral blood samples were obtained from patients at least one year after finished the chemotherapy and/or radiotherapy. Genomic alterations were evaluated using the Affymetrix high-density microarray platform (CytoScan™HD). The data were analyzed by Affymetrix Chromosome Analysis Suite (ChAS) software. The CNVs were compared with previously described CNVs and SNPs. A data set of CNVs and SNPs were present in the same platform. It was found 21 CNVs: Genomic gains on 1p31.3, 1q12q25 and 1q21.1 were detected in one case each. Nine cases presented 18 genomic losses mapped on chromosomes 1, 2, 4, 6, 7, 8, 9, 13 and 14. One female patient, who developed three malignant neoplasms (melanoma at age 40, breast at age 43 and bladder carcinoma at age 47), presented five germline losses. Interestingly, one patient had history of three tumors (breast at age 58, non-Hodgkin lymphoma at age 61 and lip carcinoma at age 65) exhibiting 17 genomic alterations. In total, 34 cases were putative candidates associated with multiple primary cancer predisposition in these families.
3233F

A patient with Angelman-like features due to deletion of chromosome 15q26.1q26.2 encompassing CHD2 and RGMa. S. Kantarcioy1, J.A. Martinez-Navarro1, 1) Department of Pathology and Laboratory Medicine, UCLA; 2) Department of Human Genetics, UCLA, Los Angeles, CA.

Angelman syndrome (AS) is a neurodevelopmental disorder caused by a variety of genetic abnormalities involving chromosome 15q11-13 region that is subject to regulation by genomic imprinting. Characteristic clinical features include severe developmental delay, speech impairment, gait ataxia, seizure, inappropriate laughter, microcephaly, strabismus, subtle dysmorphic facial features, and hyperactive lower limb deep-tendon reflexes. About 10% of individuals with clinical diagnoses of AS have an unknown etiology. Here, we report a 4.5-year-old female patient with a history of global developmental delay. The patient was born to a 39-year-old G4 P3 mother with an ectopic pregnancy at 40 weeks’ gestation. Prenatal chromosome testing was normal. At birth, the weight was 6 pounds and the length was 19 inches. She developed strabismus at 4 months of age and wears eyeglasses. By the age of 2 years, she presented speech delay, hypotonia, and increased tone on the left lower extremity. She walked at 14 months old, but toilet trained at 2.5-year-old. Although she had previously some ambulatory issues, her walking improved immediately after bilateral eye repair but she has some spasticity of the left lower extremity. She had her first words at the age of 2.5 years and sentences by 3.5-year-old. She has short stature, microcephaly, dysmorphic features (facial hypotonia with a flat midface, posteriorly rotated ears, micrognathia, and down turned corners of the mouth), and bilateral 5th finger clinodactyly. The patient was recently found to have absence seizures by EEG and they remain uncontrolled. Her parents and two siblings are healthy. Chromosomal microarray analysis (CMA) revealed an unbalanced rearrangement involving chromosomes 2 and 4 without any known etiology. Here, we report a case of 15q26.1q26.2 ranging from genomic position 93,510,242 to 96,199,396 (GRCh37/hg19). This deletion interval includes 5 RefSeq genes: CHD2 (exons 17-39), RGMa, MCTP2, LOC400456, and LOC145573. Parental CMA studies were not available. Chromosome analysis (CMA) using Affymetrix CytoHD SNP array revealed a 2.7 Mb deletion of chromosome 15q26.1q26.2 (224,577,514-225,253,174)x3, and 1q42.11q42.12(224,577,514-225,253,174)x3. We did not observe any spontaneous rearrangements in the affected chromosomes, but observed that the presence of deletions in both chromosomes 1 and 2 is associated with Angelman syndrome-like phenotype. Our report supports the significance of chromosome 15q26.1 deletion encompassing CHD2 and RGMa in Angelman syndrome-like phenotype.

3233F

Telomerase gene copy number is increased in IBD and PSC. A. Amiel1,2, Y. Sulayev1,2, H. Katz1,2, A. Stein3, M. Liberman4, F. Konikoff5, I. Laish1,2,1) Laboratory of Cytogenetics and Reproductive Biology and Human Genetics, Farhat Hached University Teaching Hospital, Sousse, Tunisia; 2) Department of Gynecology, Farhat Hached University Teaching Hospital, Sousse, Tunisia; 3) Department of Endocrinology, University Hospital Pattouma Bourgiba, Monastir, Tunisia. Telomerase is an enzyme complex that lengthens telomeres. It is composed of the catalytic subunit of human telomerase reverse transcriptase (TERT) and the telomerase RNA component (TERC) encoded by the TERC gene. Gene amplifications involving the TERC gene (3q26) are frequent in human tumors. Telomere capture (TC) is an alternative process through which broken chromosomes can acquire new telomeres. Primary sclerosing cholangitis (PSC) and inflammatory bowel disease (IBD) are pre-malignant conditions. The aim of this study was to evaluate the TERC gene copy number and TC status in peripheral lymphocytes of patients with PSC and IBD, as a possible surrogate marker for increased tendency for malignancy. Methods: By applying fluorescence in situ hybridization (FISH) to leukocytes of 14 PSC patients, 13 IBD patients (8 with Crohn’s disease, 5 with ulcerative colitis) and 12 healthy controls, we estimated gene dosage of the TERC gene at 3q26.3. We used the SNRPN and 13q14.3 genes (red) and the 13qter and 13qter (in green) to evaluate the TC phenomenon. Results: The percentage of cells with more than two copies of the TERC gene was significantly higher in PSC patients (mean 3 (± 3 red signals) 4.7) than in IBD patients (mean 2.96) and controls (mean 3.9; p-value < 0.0001). The TERC gene copy number was also higher in PSC patients with concomitant colitis (64%) despite significantly lower disease activity indices. Significantly more cells were observed with TC with both genes in IBD lymphocytes than in control cells. Conclusion: TERC gene copy number is increased in IBD and even more in PSC lymphocytes, while TC is higher in IBD lymphocytes. These findings may be related to the different tendencies and predispositions of these conditions to become malignant. In addition, TC is another mechanism for telomere elongation and is probably higher when the telomerase gene copy number is lower.

The current use of high-resolution techniques, such as genomic arrays, has revealed an unprecedented number of different cytogenomic unbalances. Among them, one of partial trisomy 19q, a rare aneusomy, had only a few cases reported in literature, with different breakpoints and an inconsistent clinical delineation. We report on a female patient with mild phenotypic findings and an important cognitive alteration, the only child of a non-consanguineous couple. At eight years old, she presents some facial dysmorphism, astigmatism, short stature, long fingers, Raynaud’s phenomenon, skeletal alterations, neuropyschomotor developmental delay and intellectual disability. She had recurrent infections, urolithiasis and seizures beginning at age four years old. Karyotype using G-banding at 550 band-resolution was normal. Subsequent studies with the MLPA (Multiplex Ligation-dependent Probe Amplification) technique, using SALSA MLPA kit P070 Human Telomere-5 containing probes for subtelomeric regions, revealed three copies of 19q. Genomic array using the Affymetric Genome-Wide Human SNP Array 6.0 showed a 10.6 Mb duplication of the long arm of chromosome 19 resulting in: arr19q13.33q13.43(48,463,121-59,097,842)x3. FISH with a BAC probe for 19q13.43 (RP-11-39587), used to determine the genomic duplication position, revealed that the duplicated region was attached to the short arm of chromosome 21. The distal region of the chromosome 19 corresponds to a genomic segment with a very high transcriptional activity, with more than 600 known genes encompassed by the duplicated region, 30 of them associated with known diseases, including epilepsy, intellectual disability and speech delay, as presented by the patient described. This rearrangement is unprecedented in literature, and its size is unexpected in view of the phenotype observed. The phenotypic consequences of duplications are complex to be studied because the chromatin context can modify the transcriptional status of trisomic segments, especially when attached near DNA repetitive regions. Therefore, localization and orientation of the extra copy may interfere in its expression levels and, consequently, in the clinical features. This report adds important data for the genotype-phenotype correlation of this de novo trisomy 19q13.33q13.43 duplication and may indicate involvement of some specific feature(s); haploinsufficiency of SIM1 and BDNF explains the correlation of this region and emphasizes the importance of a cytogenomic approach in the study of complex rearrangements. This report adds important data for the genotype-phenotype correlation and suggests involvement of some specific feature(s); haploinsufficiency of SIM1 and BDNF explains the correlation of this region and emphasizes the importance of a cytogenomic approach in the study of complex rearrangements. This report adds important data for the genotype-phenotype correlation and suggests involvement of some specific feature(s); haploinsufficiency of SIM1 and BDNF explains the correlation of this region and emphasizes the importance of a cytogenomic approach in the study of complex rearrangements.


Small supernumerary marker chromosomes (sSMC) are structurally abnormal chromosomes that cannot be thoroughly characterised by conventional G-banding analysis and are present in 0.288% of patients with developmental delay. The unstable, gene-rich 17p11.2-p12 chromosome region is associated with various structural anomalies including SMCs. To date there are relatively few cases of chromosome 17-derived sSMCs reported in the literature. Those that are recorded are of variable size, tissue distribution and level of mosaicism. Here we report a 30-year-old male with a de novo mosaic SMC(17). Our patient presented with learning difficulties, anxiety and tremors and is a known carrier of a paternally inherited balanced translocation between chromosomes 17 and 11, which is thought not to contribute to his phenotype. Conventional cytogenetic analysis revealed the presence of two cell lines. One cell line had 46 chromosomes and the t(1;11) and the second had 47 chromosomes with the t(1;11) and an additional sSMC. FISH and array-CGH studies revealed that the sSMC comprised of the proximal and proximal short arm of chromosome 17. The sSMC appeared to be present in approximately 70% of cells. Parental chromosome analysis and array-CGH indicate that the sSMC appears to have arisen de novo. Phenotypically patients with the sSMC(17) have been described with mental retardation/developmental delay, hypotonia and epilepsy; features which can be found in Puckett-Lupski Syndrome (PLS). Whole genome SNP array analysis revealed the patient's disease of other chromosome(s) that are associated with the PLS. We believe the patient's disease is associated with the PLS.

Array-CGH analysis ruled out PWS. Half of these patients were further investigated with different array platforms. We identified syndromic genomic disorders, such as deletions of 1p36 (7 cases), 2q37 (HDAC4; 5 cases), 9q19 (SIM1; 2 cases), 9p24 (3 cases), 9q34 (EHMT1; 1 case), and 17p11.2 (RAI1; 5 cases), as well as specific CNVs already linked to obesity, such as 1p21.3 (12 cases), 2p25.3 deletions and an unbalanced translocation del(8)(8;12). Additionally, we identified 22q11.2 deletions and DiGeorge syndrome 22q11.2 duplication syndrome (1 case), and 22q11.2 distal deletion syndrome (2 cases). Known genomic disorders with incomplete penetrance and variable expressivity included 15q11.2 deletion of NIPA1 (2 cases), 16p11.2 duplication of TBX1 (2 cases), 16p13 deletion of MYH11 (1 case), 17q11.2 duplication of NF1 (1 case), and 17q21.31 duplication of MAPT (1 case). Clinical variability in well-known syndromes may facilitate the identification of disease genes.

Mechanisms of concurrent deletions and duplications at 1p36. M. Gajecka1, J.A. Karolak1,2, J. Shen1, K. Wakai1, C. Golztabch1, L.G. Shafeef1, J.M. Turbitt1. 1) Institute of Human Genetics, University of Cologne and Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poland; 2) Children's Hospital Central California, Madera, CA, USA; 4) Department of Medical Genetics, Shinshu University School of Medicine, Nagano, Japan; 5) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA, USA; 6) Paw Print Genetics, Genetic Veterinary Sciences, Spokane, WA, USA.

In Monosomy 1p36, four classes of rearrangements are identified: pure terminal, interstitial deletions, unbalanced translocations, and complex rearrangements. Here we present seven cases of complex rearrangements involving terminal or interstitial deletions and duplications. To characterize the aberrations and identify mechanisms of concurrent deletion and duplication formation, rearrangements were analyzed using array CGH, SNP arrays and other molecular cytogenetics and molecular biology methods. For each individual, deletion and duplication sizes and parental origin of the rearrangements were determined. Fiber FISH analysis demonstrated either tandem or direct duplications. Rearrangement breakpoints were determined and characterized to find that for a majority of cases, the rearrangements could not be thoroughly characterised by conventional cytogenetic analysis. In the DNA sequence level examination and computational analyses will be presented. Support: Polish Ministry of Science and Higher Education, Grant NN301238836.
Posters: Cytogenetics

3241F Wolf Hirschhorn Syndrome with epibulbar dermoids in a patient with deletion 4p associated to Xp duplication. S. Bragagnolo, M. Colovati, R.S. Guilherme, M.I. Melaragno, Centro de Genética Médica, UNIFESP - EPM, São Paulo, São Paulo, Brazil.

Wolf-Hirschhorn syndrome (WHS) is a multiple malformation genomic syndrome that results from a deletion of the critical region (WHSCR) at 4p16.3 region. This deletion is autosome in 40%-60% of individuals and is caused by a de novo deletion of 4p16 and about 40%-45% have an unbalanced translocation de novo or inherited from a parent with a balanced rearrangement. WHS is characterized by typical craniofacial features consisting of broad bridge of the nose continuing to the forehead, high forehead with prominent glabella, ocular hypertelorism, epicanthus, highly arched eyebrows, short philtrum, downturned mouth, micrognathia, and poorly formed ears with pits/ tags. All affected individuals have pre- and postnatal growth deficiency and retardation, hypotonia with muscle underdevelopment and developmental delay/intellectual disability. The ocular abnormalities include strabismus, refractive errors, downsizing palpebral fissures, microphthalmos, microcornea, iria coloboma, optic nerve coloboma, ocular cyst, ptosis, glaucoma, and nystagmus. The epibulbar dermoid, hemifacial microsomia, ear malformation and vertebral anomalies are observed in patients with oculoauriculovertebral spectrum (OAVS). It has been recently suggested, based on the molecular analysis of a balanced translocation t(4;X)(p15.3;q24.1) in an OAVS patient, that abnormal expression of BAPX1 gene, located in 4p15.3, might be involved with the etiology of OAVS. We described a six month-old female, with cleft palate, pulmonary stenosis, atrial septal defect and gastroesophageal reflux. Presented with growth deficiency, developmental delay, apparent facial asymmetry, high forehead, ears tags and left microtia, cleft palate, pectus excavatum and pre-sacral pit. The ophthalmological evaluation revealed epibulbar dermoids in the right eye. Her mother, maternal aunt, grandmother and great-grandmother have ears pits and hearing impairment. The patient’s conventional G-banded karyotype was normal at 550 band resolution but genomic array (Genome-Wide Human SNP Array 6.0, Affymetrix) revealed a ~13 Mb deletion in chromosome 4 and a ~9 Mb duplication in chromosome X (arr 4p16.3;p15.3 3 (68, 345-13,569,183)x1, Xp22.33p22.31(168,551-8,907,556)x3, GRCh37/hg19), including the BAPX1 gene in 4p15.3, a possible candidate involved in OAVS pathogenesis. This is the first patient with WHS with epibulbar dermoids, hemifacial microsomia, and other phenotypic changes common to OAVS.

3242T Goldenhar syndrome and oculoauriculovertebral spectrum (GS/OAVS): Clinical and cytogenomic study. M.E.S. Colovati, S. Bragagnolo, R.S. Guiltherme, A.G. Dantas, A.B. Perez, M.I. Melaragno. Genetics Division, Department of Morphology and Genetics, Universidade de São Paulo, São Paulo, Brazil.

Goldenhar syndrome (GS) has been characterized as having an association with external ear deformities, facial asymmetry, epibulbar dermoid, and vertebral changes. Its etiology is multifactorial, familial cases have been described with autosomal recessive or dominant. Some authors characterized microtia, hemifacial microsomia and Goldenhar syndrome phenotypes a oculoauriculovertebral spectrum (OAVS). In contrast, Goldenhar syndrome is part of the OAVS, but there are changes in other organs or systems, the most common being cardiovascular (50%), central nervous system or mental impairment (5 to 10%) and renal / respiratory (5%). The accuracy of a clinical diagnosis is a challenge for clinical geneticists, since there is a heterogeneous phenotype and etiology. To date, several chromosomal abnormalities have been associated with the syndrome involving most frequently 22q and 5p. In literature, five 22q11.2 microdeletions have been described involving the same region, suggesting a possible OAVS candidate gene. In this study we analyzed 30 patients with GS/OAVS by G-banding and we found three patients with positive cytogenetic results with two apparent chromosome heteromorphisms and one inversion: 14p+, 15p+ and inv(12)(q15p24.1). These three patients were evaluated by genomic array (Genome-Wide Human SNP Array 6.0, Affymetrix) and Multiplex Ligation-Dependent Probe Amplification (MLPA kit SALSA P250). In one patient revealed a ~581 kb deletion in chromosome 22 (arr 22q11.21(20,716,923-21,297,749)x1, GRCh37/hg19). The regions deleted includes 21 genes and it is flanked by the low copy repeats LCR-B and LCR-C, not including the 1.5 Mb DiGeorge critical region (LCR-A to LCR-B). This patient presents hemifacial microsomia, retrognathia, agenesis of external auditory canal and ossicular chain with right moderate hearing loss, soft palate cleft, and thoracic hemivertebrae. His neuro-psychomotor development is normal. We present the sixth patient with OAVS and 22q11.2 microdeletion reported in literature. This study supports the previous reports indicating that patients with GS/OAVS should be screened for 22q11.2 microdeletion and other chromosomal abnormalities in order to identify genomic loci that are potentially involved in this disease pathogenesis and also to understand the phenotypic and genetic heterogeneity in this syndrome. Financial support: FAPESP, Brazil.
3244T
Compound Heterozygous Microdeletion of Chromosome 15q13.2q13.3
Region in a Child with Hypotonia and Global Developmental Delay. P.
PRASUN1, L. SIVASWAMY2, L. SCUSSEL2, M. HANKERD4, M. KRISTOF-ICE4, S. EBRAHIM2. 1) Division of Genetics and Metabolism, Children’s
Hospital of Michigan, Detroit, MI; 2) Department of Pathology, Wayne State
University School of medicine, Detroit, MI; 3) Division of Pediatric Neurology,
Children’s Hospital of Michigan, Detroit, MI; 4) Detroit Medical Center Univer-
sity Laboratories, Cytogenetics Laboratory, Detroit, MI.

Heterozygous deletion of the 15q13.3 BP4-BP5 region (OMIM #612001) is
characterized by a highly variable clinical phenotype ranging from mental
developmental delay, epilepsy, neuropsychiatric disorders, and facial dysmorphic
features to a complete absence of symptoms. Homozygous and compound heterozy-
gous microdeletions of the same region are extremely rare with only a few cases
having been reported in the literature to date. We report on the molecular
detection of compound heterozygous 15q13.2q13.3 deletions in a 23 month old
Caucasian female with global developmental delay, generalized
muscular hypotonia, and visual dysfunction. The patient was first evalu-
ated by pediatric neurologist at the age of 3 months and found to have
generalized hypotonia, no head control, no face regard or visual fixation.
MRI of the brain and routine karyotype were obtained both of which were
normal. There was concern of cortical visual failure and a detailed examination
by pediatric ophthalmologist was done. Ophthalmologic examination and
evaluation consisting of visual evoked potential (VEP) and electroretinogram
(ERG) were normal. Patient has a 6 year old brother with global develop-
dmental delay, hypotonia, and seizure disorder. He was found to have mosa-
icism for ring chromosome 20 by G-banding karyotype. Chromosomal
microarray detected a 1.28 Mb deletion on one chromosome 15q13.2q13.3
region, and a smaller sized deletion estimated to be 410 kb in size on the
second chromosome 15q13.3 homologue resulting in nullisomy for at least
the smaller deletion within the15q13 region. The larger deletion contained
7 genes: MTMR15, TRPM1, MIR211, KLF13, OTUD7A, CHRNA7, and
MTMR10. The smaller deletion contained CHRNA7 and part of OTUD7A
genes. The deletions were confirmed by FISH. Thus this patient is a com-
pound heterozygous for the 1.28 Mb deletion on one chromosome
15q13.2q13.3 and a smaller size deletion on the other chromosome 15q13.3
region. CHRNA7 is one of the critical genes in the 15q13.3 region. Haplo-
insufficiency of CHRNA7 leads to cortical visual dysfunction while
homozygous loss of CHRNA7 leads to cortical visual dysfunction while
homozygous loss of CHRNA7 contributes to visual failure by retinal dys-
function.

3245F
Deletion 12q12 - 12q13.11: Case Report with unusual neurologic mani-
festations. A.M. Zarante 1, G. Giraldo 2, J.C Prieto 1,2. 1) Instituto de Genetica
Humana, Universidad Javeriana, Bogota, Colombia; 2) Hospital La Victoria,
Secretaria Distrital de Salud, Bogota, Colombia.

Interstitial deletions involving the long arm of chromosome 12 are rare
events and few case had been reported in the literature , the deletions are characterized by : developmental delay, growth retardation, macrocephaly,
facial dysmorphism, prominent forehead, hypertelorism, downsizing pal-
pebral fissures, upturned nose, Low-set ears, pterygium colli, cardiac anom-
aly, 2-3 toe syndactyly, single palmar creases, genitourinary malformations,
anomalies , pyeloric stenosis, Ectodermal anomaly, asthma/reactive airway, ocular abnormalities and normal MRI. We present a case of 2 Colombian
girl of 8 years old. She was born at term after a normal pregnancy, non-
consanguineous parents and family history was unremarkable. The patient
presented a global developmental delay. MRI showed cerebellar vermis
hypoplasia. The findings found on the physical examination showed evidence of
macrocephaly, narrow forehead, front and back low implantation hair, bilateral
palpebral ptosis, upturned nose, flat philtrum, thin lips, dental crowding,
with winged ears, helix folding, in the limb presents shortening and clinodacty-
lis of the 5th finger, 2-3 toe syndactyly and joint hypermobility. The band R
karyotype reported: 46, XX, the Comparative Genomic Hybridization (CGH)
analysis showed deletion of 12q12-12q13.11 (2.695 MB). The parents were
negative for deletion of 12q12-12q13.11. Then it is a novo deletion. We
describe a patient with a cytogenetically 12q deletion is about 2695 MB
detected by Array CGH this deletion has been associated with develop-
dmental delay, growth retardation, facial dysmorphism. The striking thing
about this case is the cerebellar vermian hypoplasia and microcephaly.

3246T
Molecular cytogenetic techniques in investigation of suspected microdele-
tions syndrome: an experience with 330 cases. A. Halder, M. Jain,
I. Chaudhary, V. Mohan, P. Kumar. Reproductive Biology, All India Institute
of Medical Sciences, New Delhi, Delhi, India.

Background: Microdeletion syndrome is characterized by small (< 5Mb)
chromosomal deletion in which one or more genes are involved. They are
currently associated with multiple congenital anomalies. This syndrome is
the result of haploinsufficiency of genes in the critical interval. FISH, MLPA,
QFPCR and aCGH techniques are commonly used for the diagnosis. Here
this study will assess role of FISH and aCGH in the diagnosis and research
on suspected microdeletion syndrome. Method: This study was comprised of
330 cases of suspected microdeletion syndromes. There were 184 cases of
22q11.2 microdeletion, 52 cases of Williams, 47 cases of Prader Willi/
Angelman, 18 cases of Miller Dieker, 14 cases of Retinoblastoma, 5 cases of
Trichorhinophalangeal (TRP) and 10 cases of other microdeletion syn-
dromes. FISH was carried out in all using non-commercial probes. Subse-
quentley, aCGH was performed in 55 cases (40 cases of 22q11.2 and 15
cases of other microdeletion). In another 36 (mostly 22q11.2 microdeletion)
aCGH experiment is in progress. Result: FISH was confirmatory in 29 cases
(8.8%: 20 cases of 22q11.2 microdeletion, 5 cases of Prader Willi, 3 cases
of Williams and 1 cases of TRP syndrome). There were 10 cases with mosa-
icism and 19 cases with pure deletion. Microarray was picked up CNV with
or without LOH in 72.5% of cases, mostly involving several chromosomal
loci. However, aCGH was failed to pick up mosaic cases (even 45% deleted
cell lines). Clinically suspected specific deletion was detectable only in 27.3%
cases by aCGH. Variation in deletion size, break point difference as well
as other CNVs was observed. Conclusion: We conclude FISH should not
be the first method of choice for clinically suspected microdeletion syndrome
as cost, labor & time versus benefit is unjust. We think aCGH should be
first line of investigation and FISH may be used for confirmation, detecting
mosaicism, screening family members and prenatal diagnosis. However,
microarray is likely to miss mosaic cases, if deleted cell concentration is less
than 50%. Furthermore, microdeletion syndrome best fitted with genomic
disorder as several chromosomal loci are involved in CNV with or without
LOH and alteration in deletion size or breakpoint. We did not find identical
mutations in any case, thus explaining reason for phenotypic variability between
cases.

3247F
Array CGH in pediatrics diagnosis versus karyotype. Experience with
189 cases. M. Perez Sanchez, A. Moro, J.I. Barronuevo, S. Roldan, 2
A.R. Gonzalez, 2. 1) Análisis Clínicos, Hospital Virgen de las Nieves, Gra-
 nada, Granada, Spain; 2) Pediatria. Hospital Virgen de las Nieves. Granada.
Spain; 3) FIBAO. Hospital San cecilio. Granada.

Microarray-based comparative genomic hybridization (array CGH) has pro-
vided a relatively quick method to scan the genome for gains and losses of
chromosomal material with significantly higher resolution and greater
clinical yield than was previously possible. This new methodologies have led to
identification of novel genomic disorders (microduplications and microdelec-
tions) in patients with developmental delay/mental retardation and/or multiple
congenital anomalies (DD/MR/MCA), with a significant increase in diagnostic
yield. In this study we present the result of array CGH obtained in 189
children with normal karyotype but DD/MR/MCA . The NimbleGen CGX
Cytogenic Microarrays platform was performed. The results has shown
that in 41 patients (21.7 %) was detected a chromosomal deletion or duplica-
tion previously described like pathogenic copy number variants (CNVs),
were not found in the microarray analysis. Other 55 patients (29.0 %) showed
that in 7 anomalies (49 %) had occurred de novo and was classifi-
cated as pathogenic and in 11 cases (61 %) appeared to be inherited from
an unaffected parent. In a total of 48 patients (25.4 %) was possible to detect
a pathetogenic CNVs . In contrast, only in less than 2 % of cases chromosomal
alterations can be detected when karyotyping studies are performed. Several
recent studies suggest that when aCGH is performed with a finding of
an apparently normal karyotype, the diagnostic yield increases by an additional
4-16 % in those cases with clinical suspicion. In our study, we have obtained a 25.4 % of children with pathogenic
CNVs that is higher than the most published in the literature. This increase in
detection at the detection rate probably is due to the array type utilized or for different
methods in patients selection. As a conclusion, array CGH is the most
advanced method yet for assessing genomic imbalances associated with
genetic diseases. It the diagnostic capabilities of the clinical cytogenetics
laboratory and has led to the continued discovery of novel genetic
syndromes. These technology is being implemented routinely in our laboratory when a phenotype of DD/MR/MCA autism and apparently
balanced translocations are present.
Sex Chromosomal Mosaicism In The Gonads Of Patients With Disorders Of Sex Development (DSD), Akadder, SA1, Kamel, AK1, AGnathy, HM2, Desouky, N2, Mekkawy, MK1, Hennawy, A3, Makhluf, M4, Mazen, I5, 1) Human Cytogenetics, National Research Center, Cairo, Egypt; 2) Cytogenetics, Specialized Children’s Hospital, Cairo University, Cairo, Egypt; 3) Pediatric Surgical Division, Specialized Children’s Hospital, Cairo University, Cairo, Egypt; 4) Pathology, Cairo University, Cairo, Egypt; 5) Clinical Genetics, National Research Center, Cairo, Egypt.

The phenotypic variability observed in DSD patients with sex chromosomal abnormalities depends on many factors including the presence of SRY gene, the complexity of the structural rearrangement and the presence of chromosomal mosaicism, especially 45,X cell line, which can affect the threshold of SRY gene expression. The phenotypic sex strongly depends on the percentage of Y chromosome and 45,X cells in the developing gonads. This can explain differences observed between gonadal phenotype and the karyotype. We report on ten patients with variable presentations of disorders of sex development (DSD) including ambiguous genitalia, primary amenorrhea and short stature. Conventional cytogenetics studies and fluorescence in situ hybridization (FISH) technique on peripheral blood showed a non mosaic chromosomal constitution in five patients. The other five patients exhibited mosaic cell pattern associated with different types of sex chromosomal abnormalities including 45,X, isodicentric Y chromosome, X;Y translocation and ring X chromosome. FISH analysis on paraffin embedded or fresh cultured gonadal tissue specimens was done for all patients and showed mosaicism in eight of them and a single cell line in two patients. Our study demonstrates the importance of studying sex chromosome mosaicism in the gonadal tissue of patients showing a discrepancy between their karyotype and gonadal phenotype. Using FISH technique on paraffin embedded or fresh cultured gonadal tissue specimens is recommended in those patients for better understanding of the phenotypic/karyotype correlation.

3251F Directional genomic hybridization: An improved biomarker for radiation exposure, F. Ray1,2, E. Robinson3, M. Conforth1,2, J. Bedford1,2, E. Goodwin4, S. Bailey1,2. 1) Environmental and Radiological Health Sciences, Colorado State Univ, Fort Collins, CO; 2) KromaTID Inc, 320 East Vine Drive, Fort Collins, CO; 3) University of Texas Medical Branch, Galveston TX.

Cytogenetic biomarkers have been a preferred choice for retrospective estimation of radiation exposure because they are sensitive, quantifiable and relevant to biological effects of concern. The most commonly used involve the measurement of dicentrics and symmetrical translocations. Both have shortcomings that become increasingly problematic for assays carried out at long times after radiation exposure. Dicentrics in samples from peripheral blood lymphocytes decrease with time after exposure (t1/2 1-2 years), and in the case of more stable symmetrical translocations, background levels are 10-fold higher and increase with age. Another aberration type, inversions, result from exchanges within a chromosome that reverse the orientation of the broken segment. We have developed an approach based on directional genomic hybridization (DGH) that facilitates detection of inversions with a greater than 10 fold improvement in resolution over existing techniques, allowing the detection of 1 Mb or smaller inversions. Bioinformatic analysis of the genomic sequence of each chromosome when coupled with single-stranded hybridization, produced chromatid - rather than chromosome - paints. Inversions register simply as a signal switch from one sister chromatid to the other in the inverted region. Importantly, like chromosome paints, chromatid paints also reveal translocations and dicentrics. Modeling suggested that inversions should be more common than translocations after densely vs. sparsely ionizing radiation exposures. We irradiated human cells with high LET heavy ions or low LET gamma rays using chromatic painting, compared the dose-response yields for induction of inversions, translocations and dicentrics. As predicted, the slope of the dose-response curve following heavy ion irradiations was steeper, and the yields per unit dose for inversions were higher than for either translocations or dicentrics. In another application, chromatid painting of paraffin-embedded or fresh or cultured gonadal tissue specimens is recommended in those patients for better understanding of the phenotype/karyotype correlation.
3252T

The regulatory role that microRNA’s play in gene expression has made them a target of interest in many areas of research. Unfortunately, detecting miRNA has been limited to fixed or lysed cells, which presents the problem of looking at only a single time point while utilizing duplicate samples to mimic more dynamic study’s of changes in expression levels which can be observed only in live cells. The ability to monitor miRNA levels within live cells without altering gene expression or affecting cell health allows for a more biologically relevant understanding of miRNA biology. However, monitoring RNA within intact cells can prove to be challenging with current techniques due to complex or harmful sample preparation techniques or transfection reagents. Further, amplification methods can create false positives or erroneously inflate differences. Here we describe the ability to detect miRNA levels in live intact cells without the need for transfection reagents. We compare the expression level of miR-21, miR-210, and miR-155 in live cells to that of levels seen by RT-PCR. RT-PCR data is limited to a population of cells where the live cell data is captured at the single cell level. This technique utilizes gold nano-particles conjugated to duplexed oligonucleotides. In the presence of target RNA a fluorescent reporter is released from the proximity of the gold and fluorescence can be detected. In contrast to traditional RNA detection methods our technique allows for the detection within live cells allowing us to sort them based on their miRNA expression and subsequently re-use those cells for follow up experiments. This illustrates the advantage of profiling the RNA expression in live cells with the ability to further study those cells in downstream applications. It also allows for the ability to understand the expression levels of a given miRNA target across a population of cells at the single cell level using flow cytometry as the detection platform. Detecting RNA expression levels in live cancer cells with the ability to utilize the same cells in downstream testing gives researchers the ability to perform experiments which were previously thought to be impossible.

3253F
Mosaic Maternal UPD15 in a Newborn with Complex Heart Defect. D. Pickering1, W.G. Sanger2,3, R.E. Lutz4, J. Carstens1, M. Wiggins5, B.J. Dave6. 1) Human Genetics Laboratories, Munroe-Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center, Omaha, NE; 2) Genetics Medicine Department, Munroe-Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center, Omaha, NE; 3) Prader Willi syndrome (PWS) results from lack of the paternal copy of chromosome 15q11.2-13.2, most often due to either deletion 15q11.2-13.2 or maternal uniparental disomy (UPD) of chromosome 15. Microarray analysis using high-density genome-wide SNP arrays detects UPD events as well as copy number changes. Based on copy-neutral segmental or whole chromosome isodisomy detected by SNP array, subsequent methylation studies for paternal and maternal allele are typically performed to confirm the diagnosis. We report here a newborn male referred for microarray with clinical findings of complex congenital heart defect (CHD) involving unbalanced AV canal with hypoplastic arch and dysmorphic facies, small rib cage, and undescended testes. The pregnancy was complicated by intrauterine growth retardation and heart defect detected by ultrasound. Prenatal karyotyping was performed, and microarray studies confirmed a 49,000,000 base-pair segment of copy-neutral loss of heterozygosity (LOH) at chromosome 15q14-q25. Large interstitial segments of LOH without deletion and restricted to one chromosome typically indicate uniparental heterodisomy. Confirmation of the methylation studies using primer sets covering SNPRN revealed the presence of the methylated maternal band and a significantly diminished unmethylated paternal band suggestive of mosaicism. Concurrent high-resolution chromosome analysis on 20 cells was normal. Subsequent FISH studies analyzing 50 cells using probes for ACVR1 and ACVR1B, 15q telomere probes to rule out possible comcomitant maternal bands as trisomy but no evidence of trisomy 15 in peripheral blood were also normal. There was no evidence of mosaic trisomy 15 by SNP array copy number analysis. Although fibroblast cells were unavailable, based on the faint paternal band detected in karyotype studies, the 49Mb interstitial 15q segmental LOH noted by SNP microarray and the patient’s CHD, it is reasonable to assume that a second cell line containing a paternal copy does exist. Additional studies on buccal mucosa tissue are pending. This case highlights the ability to identify rare trisomy caused by parental uniparental disomy, however, ensuring confirmatory studies may be challenging. Understanding the underlying mechanisms of UPD is essential for accurate interpretation, diagnosis, and appropriate recommendations.

3254T

Copy neutral segments with allelic homozygosity (SOH) are frequently identified in cases interrogated by oligonucleotide SNP microarrays. SOH is due to parental relatedness, chromosomal recombination or rearrangements. Depending on the genomic context, it may indicate ancestral homozygosity, uniparental isodisomy (UPD), or parental consanguinity. We collected cases from consecutive specimens sent to our clinical laboratory over the past two years. The cases were reported based on the presence of SOH >10 Mb in a single region or >5 Mb in at least two regions. The percentage of the genome encompassed by SOH regions was calculated based on the total coverage of Affymetrix Cytoscan™ HD array, which is about 2,700 Mb. Classification of the cases for the degree of parental relatedness was based on the 95% confident interval value recently published (Sund et al., 2013). Of 14,575 cases analyzed by SNP arrays, 67% had a normal result, 23% had results of unclear clinical significance, and 10% had significantly copy number variations (CNVs). For each category, about 5-6% of cases had one or more reportable SOH. Of the 872 (6% of 14,575) cases with SOH, 659 (76%) cases were interpreted as arising due to identity by descent (IBD), and 213 (24%) cases were suspected or confirmed as resulting from UPD. For the cases with IBD, an estimate of the inbreeding percentage was estimated, and 5% were suspected to have first degree or closer parental relatedness, 9% second, 19% third, 16% fourth, and 51% fifth. Suspected or confirmed UPD cases were identified involving every single chromosome. Of the 16 cases with SOH from chromosome 15 alone, six cases had isoUPD (4 confirmed by methylation studies), another six were confirmed to have mixed iso- and hetero-UPD. One case had copy neutral maternal mosaic UPD15 with modified Prader-Willi phenotype. At least ten cases with SOH were confirmed to contribute to autosomal recessive disorders, such as homozygous deletions in the MCHP1 gene (microcephaly), homozygous nonsense mutation in the growth hormone receptor gene (Laron syndrome, pituitary dwarfism), and other recessive nonsense mutations in the Prader-Willi syndrome gene (Burke lymphocyte syndrome type III). This study demonstrates that the identification of SOH, in addition to CNVs, is much more frequent than previously recognized and often reflects close parental relatedness or unpalvel UPD in many cases.
Characterization of Expression Profile of the CER1 Gene and Two Regulatory Elements in Human Mesenchymal Progenitor and Neural Stem Cells.

X. Hauge, J. Sloan. Dept Biol & Physics, Kennesaw State Univ, Kennesaw, GA.

Terminal deletions of the short arm of human chromosome 9 (9p-) cause mental retardation, cardiac abnormalities, abnormal genitalia, trigonocephaly and other craniofacial abnormalities. The gene that is of most interest for understanding this syndrome is the cerberus 1 gene (CER1) which plays a role in establishing the anterior-posterior axis in vertebrates. The ortholog of CER1 in mouse binds directly to bone morphogenic proteins (BMPs) and prevent BMPs from binding to their cognate receptors, acting as a BMP inhibitor. The inhibition may slow the ossification of sutures, allowing normal brain growth and development. A deletion of the CER1 gene or of its regulatory elements could cause a premature closure of sutures, leading to trigonocephaly. Little is known regarding the CER1 expression profile during embryonic development in humans due to legal and ethic issues. We seek to investigate the expression profile of the CER1 gene during early embryonic bone development using pluripotent human mesenchymal progenitor cells. We cultured mesenchymal progenitor cells in the osteogenic differentiating medium for various time periods, and then isolated RNA from these cells. To quantify the gene activities of the CER1 and several important osteogenic markers, such as RUNX2 and BSP, quantitative polymerase chain reactions were performed. We also examined the CER1 expression profile in human neural stem cells. Finally, we examined the regulatory functions of an enhancer and a repressor in mesenchymal progenitor cells using a dual luciferase assay. These regulatory elements were cloned from the critical region of 9p- syndrome previously by us. Enhancer clone 5008 demonstrated higher activity in the human mesenchymal progenitor cells than in adult human cell lines.

Four new patients with maternal UPD20: a phenotype of isolated growth retardation and feeding difficulties.

E. J. Bhoj, I. Krantz, N. Spinner, L. Conlin, S. Mulchandani, B. Thiel, E. Zackai, I. Krantz, N. Spinner, L. Conlin, S. Mulchandani, B. Thiel. 1) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pathology & Laboratory Medicine, Children’s Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

We present four new patients with maternal uniparental disomy (UPD) for chromosome 20 and propose that maternal UPD20 is associated with a consistent phenotype characterized by prenatal onset growth retardation and feeding difficulties in the absence of dysmorphic features. Uniparental disomy of chromosome 20 is a rare diagnosis. Previously eleven patients with paternal UPD20 have been reported, which were associated with pseudohyoparathyroidism. There are additionally four reports of patients with maternal UPD, all with prenatal and or postnatal growth retardation. However, in three of these patients the phenotype was complicated by mosaic or full trisomy 20, and in those patients psychomotor retardation was also recorded. In previously reported patients, the phenotype of individuals with UPD20 and mosaic trisomy 20 was hypothesized to be a result of the mosaic trisomic cells. This report expands on the description of isolated UPD20 with four new patients identified using genome-wide SNP arrays. All four patients were referred with isolated prenatal-onset growth retardation and feeding difficulties requiring direct gastric feeds. Two of these individuals had a significant increase in length upon growth hormone supplementation. Two of the four patients had complete isodisomy suggestive of monosomy rescue or post-zygotic trisomy rescue and two had iso-hetero disomy consistent with a trisomy rescue mechanism. Chromosome 20 harbors many known and predicted imprinted genes. The phenotype related to maternal UPD20 can be a consequence of known disease-associated imprinted loci (including GNAS) or from other imprinted genes on chromosome 20. All of these patients are undergoing extensive clinical testing to further delineate the physiologic mechanisms affected by UPD20 leading to their phenotype. Given the similarity between our patients and those previously reported, we propose that the growth retardation and feeding difficulties result from maternal UPD20 rather than mosaic trisomy 20, while the trisomic cells may be the main contributor to psychomotor retardation. This study doubles the number of previously-reported patients with UPD20, allowing for improved description of the phenotype and more accurate prognostication of the course of the disease with and without interventions such as growth hormone supplementation.
3258W
Targetted Amplicon Sequencing of Tongue cancer Genome: Indian Experience. R. Rawal, S. Baghtharia, K.C. Kotnari, S.N. Shukla, A.K. Saxena. Cancer Biology, Gujarat Cancer & Research Institute, Ahmedabad, India. Cancer of tongue is one of the most common malignant cancers of the oral cavity predominantly found in India and its subcontinents. The major cause of morbidity and mortality in these patients is due to the high local invasiveness with lymph node and distant metastasis. Lack of predictive and prognostic biomarkers leading to over- or under-treatment of patients poses significant personal and socioeconomic impact. Present study was undertaken to unravel the molecular signatures in patients with tongue cancer with and without habit of tobacco consumption. Study included two patients with cancer of oral tongue (One with habit of tobacco and one without habit of tobacco). The Genomic DNA was extracted from fresh frozen tumour tissue and peripheral blood derived Mono nuclear cell fraction in both cases. The Genomic DNA samples were segregated in to 3 pools (Pool1: Tumour genomic DNA from Patient without habit; Pool2: Tumour genomic DNA from Patient with habit and Pool3: Pooled Genomic DNA from PBMCN of both the patients). Targetted amplicon sequencing was carried out on ION-TORRENT PGM following manufacturer’s instruction using 409 gene Ampliseq cancer panel 318 chip. The output data was uploaded by Torrent Variant Caller plugin and analyzed using ION REPORTER and INGENUITY independently. Sequences were compared against UCSC human genome sequence as control. Pool1 and Pool2 results were further compared with constitutional genome (Pool0). Data were filtered using frequency (35%-100%) and coverage (20X-100X) for the exonic region. The other variables used for filtering the data were zyosity, missense mutations etc. Results were plotted Chromosome wise and Venn diagrams were plotted to identify common and unique variants (genes) showing non-synonymous mutations in exonic region. There were 47 unique variants (23genes) in Pool1 and 21 variants (13 genes) in pool2 with 6 common variants (13 genes). The unique allelic variants in tobacco habits and non habits suggestive of a tobacco independent mechanism of malignant transformation in tongue tissue.

3259T
Drugs have been improved for breast cancer but there is still a need safe and more effective treatments. Variations in BRCA1, BRCA2, CDH1, PTEN, STK11 and TP53 genes increase the risk of breast cancer. Unfortunately most risk factors of breast cancer in women are not known. TSPAN8 gene (tetraspanin8) is a member of the transmembrane 4 super family. It's cytogenetic localization is at 12q21.1. Most of the coded proteins are surface proteins. These proteins mediate signal transduction in cell development, activation, growth and motility. This gene is expressed in different carcinomas and has been related to have a relation with ovarian cancer, gastric carcinoma, colorectal cancer, polycystic ovary syndrome, esophageal cancer and diabetes. It is also seen in cell migration and invasion. With up regulation the gene occupies surrounding tissues. It causes metastasis of the lungs and liver and shortens the life span of rats. Our project’s aim is to find a connection between breast cancer and the polymorphism of TSPAN8 rs7961581 C>T. Proving that the polymorphism of TSPAN8 rs7951581 has a role in breast cancer’s etiopathogenesis could lead to a means for early diagnosis of this cancer, use of the gene as a biomarker and development of a more effective treatment. For the experimental group, 50 patients diagnosed with breast cancer in 2008-2012 were selected. The DNA was isolated from the tumor tissue. DNA was isolated from blood collecte from the control group of 50 individuals with no diagnosis of breast cancer. For the TSPAN8 gene specific primers were designed and the gene was amplified using polymerase chain reaction(PCR). After checking the quality of the amplified products using agarose gel electrophoresis, the gene was cut with Hpy166II restriction enzyme which recognizes the rs7961581 polymorphism. We identified the polymorphism type, confirmed the PCR-RFLP(Restriction Fragment Lenght Polymorphism) results with DNA chain analysis and entered our data into SPSS16.0 and evaluated the relationship between the data using the chi-square test. 23(46%) individuals from the experimental group have the genotype TT(50%); CT and 2(4%) C. 14(28%) individuals from the control group have TT,25(50%) CT and 11(22%) CC. There is a recognizable statistical relation between breast cancer and TSPAN8 rs7961581 polymorphism (pearson chi-square=0.015). However a broader investigation must be done to confirm the significance of these findings.

3260F
PPM1D Mutations in Circulating White Blood Cells and the Risk of Ovarian Cancer. M.R. Akbari, P. Lepage, B. Rosen, J. McLaughlin, H. Risch, M. Minden, S.A. Narod. 1) Women’s College Research Institute, Univ Toronto, Toronto, ON, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 3) Genome Quebec Innovation Centre, McGill University, Montreal, Canada; 4) Department of Gynecologic-Oncology, Princess Margaret Hospital, Toronto, Canada; 5) Samuel Lunenfield Research Institute, University of Toronto, Toronto, Canada; 6) Department of Epidemiology and Public Health, School of Public Health, School of Medicine, Yale University, New Haven, USA; 7) Department of Medical Oncology, Princess Margaret Hospital, Toronto, Canada.
One fifth of ovarian cancer cases are estimated to be hereditary and the genes responsible for these are being found continuously and PPM1D was the last one in this series with some unique features. We used deep amplicon sequencing of the PPM1D mutation hot spot (exon 6) in white blood cells’ (WBC) DNA of a case-control group and found 20 carriers of truncating mutations among 1,295 ovarian cancer cases and one in 834 controls (OR = 13.1, 95%CI : 1.7 - 97.5, p = 0.0005). All these mutations were mosaic in WBCs and the lifetime risk of ovarian cancer among the female first-degree relatives of PPM1D mutation carriers was similar to the risk among the female first-degree relatives of non-carrier ovarian cancer patients (HR = 1.32; p = 0.78) that suggest PPM1D mutations are probably not inherited. Also the 12-year survival of the PPM1D-positive cases was lower than PPM1D-negative cases (HR = 2.0, 95%CI : 1.21 - 3.39, P < 0.007). The high odds ratio for PPM1D carriers justifies preventive oophorectomy and approximately one percent of all ovarian cancers might be prevented through screening for PPM1D mutations in healthy women regardless of their family history.

3261W
Juvenile myelomonocytic leukemia in six months old boy with breast cancer and leukemia in his Family history. M. Akouchekian. Medical Genetics, Tehran University of Medical Science, Tehran, Iran.
Juvenile myelomonocytic leukemia (JMML) classified as a rare childhood cancer and it usually occurs in children younger than 2 years old. It is known that certain medical conditions such as neurofibromatosis type 1 can make a child more likely to develop it. We investigate the chromosomal abnormality of six months old boy diagnosed JMML patient that his grandfather died from breast cancer in age of 57 and the grandfather's father died from leukemia in age of 60. Deletion 7(q) is diagnosed in this boy. Cluster of breakpoints in 7q11 to 7q36, is with two common minimal morphological zones in q22 and in q32-34. Using loss of heterozygocity (LOH) studies and YAC libraries, a 2 to 3 Mb segment in 7q22 has been designated as proximal common deleted area; the q33-34 zone is the consensual area for the distal deletion; LOH studies suggest that a specific mechanism, such as mitotic recombination in bone marrow stem cell leading to homozygozity in both granulocytes and lymphocytes, may be implicated. We also plan to run a whole genome study in this family to investigate the relationship between cancer family histories with this rare childhood cancer. The contribution between NF1 mutation and 7q deletion in JMML disease will also investigate.
3262T
Analyses of genome-wide linkage scan data among families with aggregation of breast and prostate cancer reveals evidence for linkage at 16q21-23. J. Seebe-Dimmer1, E. Langes1, K. Zühike2, K. Cooney2, 1Karmanos Cancer Inst, Detroit, MI; 2Wayne State University, Detroit, MI; 3University of North Carolina, Chapel Hill, NC; 4University of Michigan, Ann Arbor, MI.

Purpose: Epidemiologic studies have shown a co-clustering of breast and prostate cancer suggesting that there are germline variants that increase the risk of both hormonally-driven neoplasms. Mutations in BRCA1 and BRCA2 genes may explain a small portion of the observed occurrence of both breast and prostate cancer within families. The current investigation focuses on the delineation of chromosomal regions which may harbor new genes that play a role in the aggregation of breast and prostate cancer among first degree family members. Methods: A genome-wide linkage scan was conducted on 50 families participating in the University of Michigan Prostate Cancer Genetics Project. All families had at least 2 first-degree relatives diagnosed with prostate cancer and at least one female relative diagnosed with breast cancer in a first-degree relationship with one of the participating prostate cancer cases. Genome-wide multipoint nonparametric linkage analyses for the combined phenotype of breast and prostate cancer were performed using the software Merlin. Results: The strongest evidence for linkage was detected at 16q22 (LOD=3.07 at rs722579), a region previously reported to be linked to prostate cancer. This region contains several interesting candidate genes including known prostate cancer tumor suppressor genes WWOX and ATBF1, as well as BCAR1, a gene involved in a number of critical carcinogenic processes including cell migration, growth, and differentiation. Conclusions: Next generation sequencing of the genes in this region in our linked families are in progress to identify new mutations that explain clustering of prostate and breast cancer in these families and provide us new information on shared genetic pathways between these two common cancers.

3263F
Diversity of inherited damaging mutations in all breast cancer genes in three series of breast cancer patients: young onset, triple-negative and those unselected for family history, age at diagnosis or hormone receptor status. G. Bernier1, J. Mandell2, T. Walsh3, S. Casadei4, M. Lee5, E. Swisher6, M.C. King7, 1University of Washington, Division of Medical Genetics, Seattle, WA; 2University of Washington, Department of Obstetrics and Gynecology, Seattle, WA.

In addition to BRCA1 and BRCA2, inherited mutations in multiple other genes predispose to breast cancer. Many of these genes are in the BRCA1-Fanconi Anemia complex. We used BROCA, a targeted genomic capture and massively parallel sequencing approach, to identify all classes of mutations in all known breast cancer genes. Only truncations, complete gene deletions, splice mutations known to lead to a mutant message, and missense demonstrated functionally to be damaging were counted. Study participants were three series of breast cancer patients: those sequentially enrolled selected for age at diagnosis and family history or hormone receptor status (SBC; n=533); and those unselected for family history, age at diagnosis or hormone receptor status. Of the YBC patients, 7.7% (26/344) carried mutations in genes other than BRCA1 and BRCA2; 8 in PALB2, 7 in BARD1, 3 in CHEK2, 2 in CDKN2A, 1 each in ATM, BARD1, BAP1, and NBR1. Of the YBC patients, 7.7% (26/344) carried damaging mutations in genes other than BRCA1 and BRCA2; 8 in PALB2, 7 in BARD1, 3 in CHEK2 and RAD51D, 2 each in BRIP1 and RAD51C, and 1 in ATM. The mutational spectrum of these patients was highly heterogeneous, with 53 different mutations in 13 different genes.

3264W
Germline Copy Number Variants as Genetic Risk Factors for Familial Colorectal Cancer Type X. D.D. Buchanan1, M. Clendenning1, C. Rose2, M.D. Winton3, A.K. Wigg4, J.L. Hoggett5, H.A. Jenkins6, Colon Cancer Family Registry. 1) Cancer & Population Studies Group, QIMR, Brisbane, Queensland, Australia; 2) School of Medicine, University of Queensland, Brisbane, Queensland, Australia; 3) Envi Pathology, Brisbane, Queensland, Australia; 4) Sullivan and Nicolaides Pathology, Brisbane, Queensland, Australia; 5) New Zealand Familial Gastrointestinal Cancer Service, Auckland Hospital, Auckland, New Zealand; 6) Department of Gastroenterology, Middlemore Hospital, Auckland, New Zealand; 7) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Parkville, Victoria, Australia.

Background: Only one-third of the familial risk of colorectal cancer (CRC) is explained by variants in known CRC susceptibility genes. Multiple-case CRC families that fulfill the Amsterdam-I criteria for Lynch syndrome but do not carry mismatch repair (MMR) gene mutations and demonstrate no evidence of MMR-deficiency in their tumors are referred to as meeting the Familial Colorectal Cancer Type X (FCCTX) criteria. Copy number variants (CNVs; deletions or duplications of DNA segments), have been detected in 5-10% of familial CRC patients. However, the extent to which CNVs contribute to familial CRC risk is largely unknown.

Purpose: To determine the role of CNVs in families meeting FCCTX criteria.

Methods: We identified 344 families meeting FCCTX criteria with 168 individuals with CRC (114 index cases and 54 relatives). At least 3 members of each family were genotyped with the Illumina Genome-Wide SNP Array 6.0. CNVs were defined as oligo homozygous segments using the GenomeStudio software (Illumina). CNVs were classified as deletions or duplications based on read depth analysis. The number of CNVs that were present in all CRC-affected individuals in a family but not detected in the tests or seen in the DGV ranged from 5 to 13 deletions per family and 14 to 30 duplications per family. Candidate CNVs included a duplication in the high-copy number tumor suppressor FHIT gene (mean age at diagnosis =56.7 years ± standard deviation (SD)= 11.7 years), 3 early-onset poly-affect (mean age =45.6 years ± SD= 11.3 years) and 15 unaffected family members of 15 FCCTX families (minimum 2 CRC-affected and 1 unaffected relative/individual). Poly-duplicated. The number of CNVs that were present in all CRC-affected individuals in a family but not detected in the controls or seen in the DGV ranged from 5 to 13 deletions per family and 14 to 30 duplications per family. Candidate CNVs included a duplication in the high-copy number tumor suppressor FHIT gene (mean age at diagnosis =56.7 years ± standard deviation (SD)= 11.7 years), 3 early-onset poly-affect (mean age =45.6 years ± SD= 11.3 years) and 15 unaffected family members of 15 FCCTX families (minimum 2 CRC-affected and 1 unaffected relative/individual). Poly-duplicated.

Results: The significance of CNVs in families meeting FCCTX criteria remains poorly understood. The aim of this study was to identify rare or novel germline CNVs that predispose to CRC in families that meet the FCCTX criteria. Methods: Blood-derived DNA from CRC- and poly-affect and unaffected relatives from 16 multiple-case CRC families fulfilling the FCCTX criteria and from 16 unrelated controls from the Australasian Colorectal Cancer Family Registry were genotyped using the Illumina HumanSNP6.0 BeadChip. CNVs in FCCTX families were identified using Partek GS v6.6, a software package that allows detection of CNVs from input of read depth data. Candidate CNVs were identified that encompassed genes, segregated with CRC, were absent in unaffected relatives and unrelated controls, and either not previously reported in the Database of Genomic Variants (DGV) or were reported at rare variants in the DGV. Results: Forty CNVs were detected in FRCC families and 21 CNVs were detected in the controls. Candidate CNVs were identified that encompassed genes, segregated with CRC, were absent in unaffected relatives and unrelated controls, and either not previously reported in the Database of Genomic Variants (DGV) or were reported at rare variants in the DGV. The number of CNVs that were present in all CRC-affected individuals in a family but not detected in the controls or seen in the DGV ranged from 5 to 13 deletions per family and 14 to 30 duplications per family. Candidate CNVs included a duplication in the high-copy number tumor suppressor FHIT gene (mean age at diagnosis =56.7 years ± standard deviation (SD)= 11.7 years), 3 early-onset poly-affect (mean age =45.6 years ± SD= 11.3 years) and 15 unaffected family members of 15 FCCTX families (minimum 2 CRC-affected and 1 unaffected relative/individual). Poly-duplicated. The number of CNVs that were present in all CRC-affected individuals in a family but not detected in the controls or seen in the DGV ranged from 5 to 13 deletions per family and 14 to 30 duplications per family. Candidate CNVs included a duplication in the high-copy number tumor suppressor FHIT gene (mean age at diagnosis =56.7 years ± standard deviation (SD)= 11.7 years), 3 early-onset poly-affect (mean age =45.6 years ± SD= 11.3 years) and 15 unaffected family members of 15 FCCTX families (minimum 2 CRC-affected and 1 unaffected relative/individual). Poly-duplicated.

Conclusions: We have identified several rare or novel germline CNVs that segregate with CRC-affected members in families that meet the FCCTX criteria. Future studies will extend the study to genotype additional FCCTX families, test a large series of CRC cases and controls for variants in these genes and further characterise the effects of these candidate CNVs on gene expression.

3265T
HLA-DPB1 polymorphisms and cervical squamous cell carcinoma risk. T. Chang1, Y. Yang2,3, Y. Lee4,5,6, T. Chen6, W. Lin7, S. Chang1, 1Med Res Dept, Mackay Memorial Hosp, New Taipei, Taiwan; 2Department of Gynecology and Obstetrics, Mackay Memorial Hospital, Taipei, Taiwan; 3Department of Gynecology and Obstetrics, Taipei Medical University, Taipei, Taiwan; 4Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan; 5Pediatrics, Taipei Medical University, Taipei, Taiwan.

Cervical cancer is a multifactorial disease and infection by oncogenic human papillomaviruses (HPVs) represents the major environmental risk factor. However, the mere presence of HPVs is not enough for cervical cancer development and host immunogenic background may play an important role as well. Variations in human major histocompatibility genes may alter the efficiency of immune response to HPV antigens and have been implicated in the risk for cervical cancer. The aim of this study is to examine the role of human leukocyte antigen (HLA)-DPB1 gene in cervical cancer susceptibility. We used high-resolution methods to genotype HLA-DPB1 in 344 cervical squamous cell carcinoma (CSCC) and 350 age/sex matched healthy controls. The presence and genotypes of HPV in CSCC patients were determined by PCR. We found the HLA-DPB1*13:01 was associated with an increased risk of CSCC (OR = 1.68, 95% CI 1.17-2.43) and HPV-16 positive CSCC (OR = 1.94, 95% CI 1.26-2.98). However, the significance remained only in HPV-16 positive CSCC patients after Bonferroni correction (IterativeTextPSubscript = 0.02). In conclusion, our results suggest that HLA-DPB1*13:01 is involved in the genetic susceptibility to HPV-16 positive CSCC in the Taiwanese population. Further studies, with larger cohort are required to confirm the role of DPB1*13:01 in the development of CSCC.
3266F
Replication of cervical cancer susceptibility loci identified in genome-wide association study in a northern Swedish population. D. Chen1, J. Hammar1, D. Lindquist2, U. Gyllensten1. 1) Department of Immunology, Genetics and Pathology, Uppsala University, Sweden; Uppsala, Sweden; 2) Department of Radiation Science, Umeå University, Sweden.
In a genome-wide association study (GWAS) we have previously identified and performed the initial replication of three novel susceptibility loci for cervical cancer: rs9272143 upstream of HLA-DRB1, rs2516448 adjacent to MHC class I polypeptide-related sequence A gene (MICA) and rs3117027 at HLA-DPB2. The risk allele T of rs2516448 is in perfect linkage disequilibrium with a frameshift mutation (A5.1) in MICA exon 5, which results in a truncated protein. To validate these associations in an independent study and evaluate effect modification by age of onset and tumor stage, we genotyped the single-nucleotide polymorphisms at rs2516448, rs9272143, rs3117027 and the MICA exon 5 microsatellite in 961 cervical cancer patients and 1725 cancer-free control subjects from northern Sweden. Association between each variant and cervical cancer risk was estimated by logistic regression analysis. The C allele of rs9272143 conferred protection against cervical cancer (odds ratio [OR]=0.73, 95% CI=0.65-0.82; P=0.001), whereas the T allele of rs2516448 increased the susceptibility to cervical cancer (OR=1.36, 95% CI=1.21-1.52; P=1.8×10−7), with the same association shown with MICA-A5.1. We also identified protective effects of the MICA-A4 (OR=0.79, 95% CI=0.67-0.93; P=4.1×10−3) and MICA-A5 (OR=0.60; 95% CI=0.50-0.72; P=5.2×10−5) alleles. The direction and the magnitude of these associations were consistent with our previous findings. None of the variants studied showed heterogeneity by age. No association was observed between rs3117027 and risk of cervical cancer (OR=0.99, 95% CI=0.87-1.11 for allele A; P=0.80). Our results support the role of HLA-DRB1 and MICA in the pathogenesis of cervical cancer.

3267W
Laser Capture Microdissection (LCM) is a technique providing a rapid and reliable method to procure purified cell populations, either as multiple single cells or a group of cells, from a heterogeneous tissue sample allowing a targeted approach to genomic profiling. When analyzing gene expression profiles from a group of cells, the average profile may not be a true representation of the many different profiles that could exist even in a pure cell population (e.g., in different states of growth, differentiation, or activation). As a result, the transcriptional variability of individual cells and any insight into the relationship between specific genes in single cells are lost. To fully understand the complexity of tissue and cellular heterogeneity, it is necessary to measure molecular signatures at the single cell resolution. Here we developed a workflow for gene expression profiling of a single LCM cell that can also be applied to small number of LCM cells. The ArcturusXT™ LCM system was used to harvest single cells from frozen human tumor tissue samples which were then processed and analyzed with a quantitative RT-PCR assay platform.

3268T
Loss in the function and homeostasis of the PDE family might inflict on the cAMP and cGMP signaling and lead to the development of prostate cancer. R.B. de Alexandre1,2, A. Horvath1, A.D. Manning1, N. Halipoglou1, F. Kardauke1, D. Carraro1, F. Soares3, M. Nesterova1, C. Stratakis1, F.R. Faučz1,2. 1) Section on Endocrinology & Genetics, PDEGEN, NICHD, NIH, Bethesda, MD, USA; 2) Laboratory of Molecular Genetics, NIMA, PPGCS, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil; 3) Laboratory of Genomics and Molecular Biology - A.C. Camargo Hospital, São Paulo, Brazil.
The phosphodiesterases (PDEs) are a family of 11 intracellular isoforms coded by 21 different genes. They are responsible for hydrolyzing cAMP and cGMP to their respective 5’-nucleoside monophosphate. Some PDEs are specific for the hydrolysis of cAMP or cGMP, while others have mixed specificity. Each PDEs has a different ability to control each of these messengers concentration in different tissues, provoking different cellular responses and functions. Cyclic AMP and cGMP intracellular levels depend on the balance between their synthesis and degradation. Variants in PDEs have been associated with different disorders. The first PDE to be implicated in the predisposition of prostate cancer (PCa) was the PDE11A, whose inactivating mutations have been reported to be frequent in its patients. Further, higher levels of cAMP have been measured in prostate cancer compared to the normal tissue. In addition, cAMP and cGMP signaling is involved in PCs cell growth and modifies androgen receptor effects. We sequenced 16 different PCa tumor DNAs on SOLiD4 platform, after targeted enrichment for the coding parts of 196 genes, including the family of the phosphodiesterases, adenylyl and guanylate synthases, all genes involved in cAMP and cGMP pathways and other genes related to endocrine tumors. After alignment and variation calling, Sanger sequencing confirmation and biostatistics analysis (through comparison patients with controls of the 1000 Genome Project), we were able to select 22 different SNPs associated with the disease. These significant SNPs were found to be located in the PDE1C, PDE2A, PDE4B, PDE5A, PDE6A, PDE6B, PDE8C, PDE8A, PDE8B and PDE11A genes. Moreover, it was found two novel mutations confirmed by Sanger located in the PDE1A and PDE7B genes. Furthermore, the previously identified are still seven novel mutations with large chances to be real positives among all the PDEs waiting to be confirmed by Sanger sequencing. These results demonstrate that there might be a set of disarrangements in the intracellular levels of cyclic AMP and cGMP altering its balance. This data also establish a link between the PDEs and PCa, suggesting that, not only what was report previously with only the PDE11A, but that the whole family might be involved with a higher susceptibility to PCs. However, further investigations at protein level, trying to identify what exactly each alteration might influence in the final phenotype, are needed to be accomplished previously publication.
3269F
Genetic variants in miR-499a, miR-938 and miR-1206 are associated with gastric cancer in Europeans from the EPIC-EURGAST study. Y. Espinosa-Parrilla1,2, X. Muñoz3,4, I. Torquella-Loran1, C. Bone1, N. García5, E. Riboli6, C.A. González5, N. Sala6, I.T. Torruella-Loran5, C. Bonet5, E. Riboli6. 1) CIenciencias Experimentales y de la Salud (CEX-UPF), Institutt de Biologia Evolutiva (UPF-CSIC), Barcelona, Barcelona, Spain; 2) Programa de Genética Humana, ICBM, Facultat de Medicina, Universitat de Chile, Santiago de Chile; Chile; 3) Hereditary Cancer Program; Catalán Institute of Oncology (ICO-IDIBELL), Barcelona, Spain; 4) Molecular Epidemiology Group, Translational Research Laboratory, Catalán Institute of Oncology (ICO-IDIBELL), Barcelona, Spain; 5) Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, ICO-IDIBELL, Barcelona, Spain; 6) Department of Epidemiology and Public Health, Imperial College London, London, United Kingdom.

MicroRNAs (miRNAs) are post-transcriptional gene regulators that participate in diverse biological pathways and may act as either tumor suppressors or oncogenes. Single nucleotide polymorphisms (SNPs) in miRNAs may contribute to cancer development causing changes in either miRNA expression or function. To look for a possible contribution of miRNA genetic variants to gastric cancer (GC) susceptibility we selected 40 SNPs potentially functional because of their location in the seed (4), mature (5) or precursor (31) sequence of 40 miRNAs and genotyped them in 1284 controls matched to 365 incident GC cases with different histological and tumor location phenotypes (European Prospective Investigation into Cancer and Nutrition (EPIC) cohort). Logistic regression analysis under the log-additive model showed that rs21143587/T/C, located in the precursor sequence of miR-1206, was associated with the non-cardias localization of the adenocarcinoma (p-value= 0.0093, OR (95% CI)= 0.73 (0.58-0.93)); this SNP could affect the dosage of the corresponding miRNA by altering its biogenesis or stability. Furthermore, two out of the four SNPs located in miRNA seed regions were associated with GC: rs12416605C/T in miR-938 associated with the diffuse localization of the adenocarcinoma (p-value= 0.0281, OR (95% CI)= 0.70 (0.51-0.97)) and rs3746444/C in miR-499a-3p associated with the diffuse phenotype (p-value= 0.0308, OR (95% CI)= 0.64 (0.42-0.98)). To investigate if differential gene regulation could be underlying the genetic association of GC with these SNPs in miR-938 and miR-499a we predicted target genes for the different miRNA alleles of both variants using TargetScan. In total we found 75 predicted genes for the different alleles (three common target genes for the two alleles of rs12416605 and no common target genes for the two alleles of rs3746444) as well as a significant decrease in the number of genes predicted to be regulated by the minor alleles of rs12416605 (115 and 75 predicted target genes for the C and T alleles, respectively) and rs3746444 (259 and 9 predicted target genes for the T and C alleles, respectively). These results indicate that specific miRNA allele variants are associated with GC susceptibility in European populations probably influencing either miRNA dosage or the number and spectrum of target genes regulated by these miRNAs.

3270W
Bladder cancer susceptibility variants within CCNE1 are associated with miRNA expression of an alternative splicing form. Y.P. Fu1, I. Kohaar1, W. Tang1, P. Porta-Car Europeans from a genome-wide association study (GWAS). CCNE1 regulates cell cycle and thus is a strong candidate gene for cancer susceptibility. METHODS: Based on the bladder GWAS and the 1000 Genomes Project reference panel (phase 1 version 3.2, 2012 March revised), we imputed 4,650 SNPs within 200 K of the region (chr19:30,102,901-30,515,215) among 3,532 bladder cancer cases and 5,120 controls of European ancestry. CCNE1 mRNA expression was evaluated by RNA-Sequencing in 6 pairs of tumor-normal bladder tissue samples and with custom-designed TaqMan assays in 42 muscle-invasive bladder tumor and 42 adjacent normal tissue samples. Logistic regression and multivariable linear models were used to test for the association between SNPs and bladder cancer risk, and between SNPs and miRNA expression. All models were adjusted for age, gender, smoking habits when applicable. RESULTS: A total of 700 well-imputed and 76 GWAS-genotyped SNPs in the CCNE1 region were used for analyses. We identified an imputed SNP rs7257694, a coding synonymous variant (Ser390Ser) in CCNE1 gene, was in strong linkage disequilibrium with rs8102137 (r2=0.91, p=2.7e-160). Through actual genotyping, rs7257694 showed a per-allele odds ratio (OR) of 1.10 (95%CI=1.03-1.17) for bladder cancer risk, which was comparable to rs8102137 (OR=1.12, 95%CI=1.05-1.20) in the same samples. In addition to the main CCNE1 transcript, RNA-sequencing revealed 2 splicing forms of CCNE1 with deletion of exons 5 or 7. TaqMan assays showed significantly higher mRNA expression in bladder tumors than in adjacent normal bladder tissues of total CCNE1 (p<3.97E-12) and of a transcript without exon 5 (p=6.15E-10). The mRNA expression of CCNE1 transcript without exon 7 was significantly associated with both rs8102137 and rs7257694 in normal and tumor bladder tissues (all p<0.05). CONCLUSION: In addition to the GWAS variant rs8102137, we identified a new synonymous SNP rs7257694 as a variant associated with bladder cancer risk and with mRNA expression of the CCNE1 transcript without exon 7. Future studies are warranted to validate our findings in additional samples and investigate the possible functional mechanisms of this genetic association in bladder cancer.
Few mutations in known high-risk gastric cancer susceptibility genes in Chinese gastric cancer kindreds. A.M. Goldstein1, N. Hu1, L.-J. He2, X.-Y. Han1, J. Ho1,4, M. Rotunno1, M. Malasky1,4, H. Su1, L. Wang1, C. Wang1, L. Burdett1,4, B. Hicks1,4, K. Jones1,4, J. Boland1,4, A. Hutchinson1,4, M. Yeager1,4, T. Ding3, C. Gillen2, M.A. Tucker1, S.J. Chanock1, M. Lee6, P.R. Taylor1. 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, USA; 2) YangCheng Cancer Hospital, YangCheng, Shanxi, PR China; 3) Shanshi Cancer Hospital, Taiyuan, Shanxi, PR China; 4) Cancer Genetics Research Laboratory, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA; 5) Information Management Services, Inc., Silver Spring, MD, USA; 6) Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland, USA.

Gastric cancer (GC) causes more than 700,000 deaths each year. Familial GC results from the complex interplay of genetic and environmental factors. Several high-risk susceptibility genes for familial GC and related disorders have been identified, but most of these genes have not been examined in families from high-incidence regions such as the Taihang Mountains of North-Central China, a region with some of the highest rates reported for GC and esophageal cancer (EC). We searched the Human Gene Mutation Database (HGMD) and COSMIC to select germline disease-causing genes in familial GC. To allow for related phenotypes, we also included susceptibility genes for gastrointestinal polyps and colorectal cancer. Fourteen genes (CDH1, MET, MUTYH, TP53, MLH1, MSH2, MSH6, PMS1, PMS2, APC, CTNNB1, SMAD4, BMPR1A, MADH4) were selected. Families with 3 or more GC and/or EC cases (including ≥2 GC patients) from an ongoing family study were eligible for this study. Twelve eligible families were selected for exome sequencing; 8 families had at least 3 GC cases. The goal of our study was to determine whether high-risk GC-related susceptibility genes are important in GC families from the Taihang Mountains. Twenty-six patients with available DNA were exome sequenced. We interrogated the exome sequencing data for rare co-segregating mutations in the 14 GC-related susceptibility genes. Rare co-segregating variants were found in only 3 (CDH1, MSH2, MUTYH) of the 14 genes in two GC families. An MSH2 substitution (rs63750716) was found in 2 patients (1 GC, 1 EC) with available DNA from a case family (2 GC; 2 EC); it was seen at 0.3% frequency in Asian 1000 Genomes subjects and predicted to be neutral (MutationAssessor). Variants in CDH1 and MUTYH were seen in a single family with 3 GC patients, but, each variant was seen in only 2/3 GC cases.

Acute lymphoblastic leukemia (ALL) occurring in the first year of life is rare, accounting for 2-5% of pediatric ALL cases. Infant ALL is distinguished by unique clinical and biological characteristics, with an aggressive course following a short latency period. The mixed lineage leukemia (MLL) gene, located on chromosome 11q23, is involved in 80% of cases. Currently, 79 different MLL-fusion partner genes have been molecularly characterized with t(4;11), t(9;11) and t(11;19) the most frequent translocations in infant ALL. In this study we focused on MLL-rearranged infant ALL where diagnosis occurred at <92 days. At present, the outcome for these infants remains poor with 26% five-year survival. Given the advent of next generation sequencing, further insight into the biology of the disease may identify potential targets for novel therapies and ultimately improve outcome. We performed RNA-sequencing (Illumina, 100bp paired end) on six primary patient infant ALL samples: three patients had the t(4;11) translocation, one patient had a t(11;19) translocation and a pair of monozygotic twins with a rare MLL-translocation partner gene, t(1;11). Upon alignment of sequence reads to reference sequences (including genome, splice junction and transcriptome sequences), a pipeline utilizing Genome Analysis Toolkit (GATK) functions, Annovar, SIFT, Polyphen2, dbSNP and COSMIC was used to process the sequence alignments and annotate single nucleotide variants (SNVs). Ingenuity Pathway Analysis was performed on gene lists associated with predicted damaging SNVs. This revealed an over-representation of cancer-associated genes harboring damaging SNVs that were shared among all six infant ALL patient samples. This gene set involves multiple genes previously reported to be involved with hematological neoplasia that may represent novel therapeutic targets for treatment. We also sought gene fusions in these datasets using FusionFinder that led to the identification of a number of novel putative gene fusions involving known oncogenes. Further studies are required to examine the role of these SNVs and gene fusions in infant leukemogenesis.
3273W
Chemotherapy-Induced Peripheral Neuropathy and Cognitive Dysfunction: Role of Genetic Variation. K. Holohan1,2, Y. Wang2, B.C. McDon-ald1,2, S. Corrigan1, D.J. Smith1, J.D. West1, K. Nho3, S. Kim3, A.J. Saykin1,3,4. 1) Medical and Molecular Genetics, Indiana University-Purdue University Indianapolis, Indianapolis, IN; 2) Training in Research for Behavioral Oncology and Cancer Control Program, Indiana University School of Nursing, Indianapolis, Indiana; 3) Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, Indiana; 4) Indiana University Melvin and Bren Simon Cancer Center, Indiana University School of Medicine, Indianapolis, Indiana.

Although the relationship of pain and memory in many disease models, this topic has not yet been investigated with chemotherap-apy-associated peripheral neuropathy and cognitive dysfunction. Previous studies have reported cognitive dysfunction measured by decreased neuropsychological test performance, compared to baseline, as well as alterations in neuroimaging measures, and treatment with paclitaxel and docetaxel has been strongly associated with increased incidence and severity of peripheral neuropathy symptoms (PNS). Significant genetic modifiers have been identified for both of these adverse events, but possible correla-tions between them have not been explored. We hypothesized that chemother-apy-induced peripheral neuropathy and cognitive dysfunction may be associated, and that these measures may be modified by genes EPHA5, FGFD4, or FZD3, which have been previously associated with chemotherapy-induced peripheral neuropathy. We examined this hypothesis in a prospect-ive cohort of breast cancer patients treated with (Ctx-, n=27) and without (Ctx+, n=26) chemotherapy and healthy controls (HC, n=26) studied at baseline (BL: post-surgery, before chemotherapy) and one month after treatment completion. DNA was extracted from peripheral blood mononuclear cells and spin labeling MRI to obtain cerebral blood flow (CBF), cognitive complaints were assessed using the Multiple Ability Self-Report Questionnaire (MASQ), PNS were assessed using the 9-item FACT/GOG-Ntx subscale, and all subjects were evaluated using the Functional Assessment of Cancer Therapy (FACT) questionnaires. The neuroimaging analysis of Ctx+ patients using SPMI indicated that PNS at 1M covarying for BL were positively associated with CBF change from BL to 1M in the left anterior cingulate region (p<0.001 uncorrected), which has been associated with pain processing. This hypothesis of CBF change was set-tested in PLINK for association with EPHA5, FGFD4, and FZD3. The EPHA5 gene set was significantly associated with CBF change, as well as with PNS at 1M covarying for BL (p<0.03). Further analysis indicated that one of the most significant SNPs (rs13149846), was also associated with the language subsdomain of the MASQ (p=0.028), suggesting that this gene may be involved in both pain processing and cognition, and could be a potential target to address both of these issues. Future investigation should test EPHA5 for association with other measures of cognitive processing and replication in other cohorts.

3274T
Occult hepatitis B and carcinogenic markers in chronic hepatitis C infection. R. Rose. molecular diagnostics, institute of genetic engineering, Cairo, Egypt.

The presence of HBV-DNA in the patient’s serum without detectable HBV surface antigen (HBsAg) called occult infection. Detection of occult HBV infection in chronic hepatitis C virus patients was investigated by using qualitative PCR. Co-infection with occult HBV in chronic HCV patients increases the risk for progression to hepatocellular carcinoma (HCC). Detection of CD45-C90+ as a biomarker in HCC patients by flow cytometry. We searched for serum HBV DNA in 30 patients with histologically verified HCV-related chronic liver disease, in addition to 10 healthy control subjects collected at National Liver Institute in Shebin El-Korn, Monofiya University, Egypt from January, 2010 to October, 2010. Off 40 patients, the sera of 9 (15.0%) were positive for HBV DNA by the different PCR assays, document-ing an occult HBV infection. It found that 5 patient samples are positive for HBV DNA (Surface gene) (12.5%) of total 40 patient samples, also 3 patient samples are positive for HBV DNA (X gene) (7.5%) of total 40 patient samples, and only one patient sample is positive for HBV DNA (core gene) (2.5%) of total 40 patient samples. Only two samples from the nine positive samples were positive for both X-gene and Surface gene. In conclusion these data suggest that occult HBV infection may have clinical signifi-cance in chronic hepatitis C patients. Keywords: Biomarker, Flowcytometry, Hepatocellular carcinoma, Occult HBV, Introduction Occult hepatitis B infec-tion (OBI) has been recognized for nearly 20 years. However, with the improvements in sensitivity of serological and genomic amplification assays, the frequency and attention given to this nearly silent form of the infection has been increasing. Because low or very low viral load is part of the definition of OBI, molecular data are scarce, and, because most patients affected are asymptomatic, little clinical data, in particular histological, is available (Allain J.P. 2005). Hepatocellular carcinoma (HCC) is one of the most common malignant tumors worldwide ranging between 3% and 9% annually. Vel azquez et al., 2003. In Egypt, HCC reports to account for about 4.7% of chronic liver disease patients. HBV and HCV infections are strongly associ-ated with liver cirrhosis and HCC (Rahman et al., 2001. Africa ., 2008).

3275F
Fine mapping of the Finnish hereditary prostate cancer linked loci at 2q37 and 17q12-q22. V.H. Laitinen1, T. Rantapero1, D. Fischer4, E.M. Holohan1, T.L.L. Tammela2, T. Wahlströ1, J. Schleutker2,4. 1) Institute of Biomedical Technology/BioMedTech, University of Tampere and Finlab Laboratories, Tampere, Finland; 2) School of Health Sciences, University of Tampere, Tampere, Finland; 3) Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Tampere, Finland; 4) Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Turku, Finland.

Linkage studies of Finnish hereditary prostate cancer families have revealed a strong connection between prostate cancer and two chromo-somal regions, 2q37 and 17q12-q22. The 17q12-q22 locus is the same previously reported in the US population from Michigan and recently, both of these genomic regions have been detected in multinational GWAS analyses. Although the recently identified GBA4 variant in the HOXB13 gene at 17q21- q22 has been detected at a high frequency in Finnish prostate cancer patients, this variant alone does not explain the observed linkage to 17q12- q22. Neither is the candidate gene at 2q37 known for the present. To explain the linkage, we therefore screened these two regions of interest by next-generation sequencing (NGS). The NGS produced over 100,000 unique sequence variants which were carefully filtered in a multistep prioritization process, yielding a subset of 58 putative prostate cancer associated variants co-segregating with the disease in the analyzed families. These variants were then validated in 1293 Finnish prostate cancer cases and in 923 controls with Sequenom MassARRAY system. Statistical analyses revealed 13 SNPs in 7 genes that were significantly associated with prostate cancer. Two novel susceptibility alleles were identified in the ZNF652 gene (17q21.3). In addition, we found a deleterious frameshift mutation in the EFCAB13 gene (17q21.3). As a complementary approach, we investigated the impact of prostate cancer associated sequence variants on the regulation of genes located within the two linked regions. This targeted expression quantitative trait level analysis was performed by combining the NGS data with 4) transcriptome data obtained from RNA sequencing. In total, five candidate eQTLs were found at 2q37 and 181 candidates at 17q12-q22 before multiple testing adjustment. The novel prostate cancer associated SNPs identified in the NGS analysis will be further studied, and a possible support the suggested role of ZNF652 as a prostate cancer candidate gene. Moreover, this is the first targeted resequencing project studying the locus on 2q. The regulatory elements discovered by eQTL mapping provide new insights into the complex genetic events contributing to prostate cancer predisposition.

3276W
Novel EZRERBB4 fusion gene found in follicular variant of papillary thyroid cancer. S. Lee1, H.-G. Jee1, K. Lee2, J.-S. Seo3, K. Holohan1,4. 1) Biomedical Science, Seoul National Univ, Seoul, South Korea; 2) Seoul National University Hospital, Seoul, South Korea; 3) Department of Neurosurgery, Seoul National University; 4) Medical Oncology and Cancer Control Program, Indiana University School of Medicine, Indianapolis, Indiana; 5) Department of Radiation Oncology, Tampere University Hospital, Tampere University Hospital, Tampere, Finland.

Thyroid cancer is one of the most prevalent cancers worldwide, and papillary thyroid cancer accounts for ~80% of all thyroid cancers. Here we conducted RNA sequencing for 46 papillary thyroid cancer tissues in Korean patients and explored any fusion mutations found using the TopHat-Fusion program. As a result, we discovered 3 fusion mutations using our stringent criteria and these include CCDC6/RET and PAX8/PPARG fusions, which have been previously reported for this cancer type. The other EZR/ERBB4 fusion is a novel fusion mutation found in papillary thyroid cancer, and the sample having this mutation was diagnosed as the follicular variant subtype with abnormal pathological findings. This is an in-frame mutation and the expression pattern verified this fusion at the exon level. Previous reports have shown that ERBB4 is a driver gene for carcinogenesis in multiple types of cancers. In this study, we suggest the EZR/ERBB4 fusion as a novel driver mutation in papillary thyroid cancer, showing the atypical follicular patterns of pathology.
3277T

Sequence variants in \textit{BARD1} and breast cancer susceptibility: results from the Breast Cancer Family Registry study. F. Lesueur\textsubscript{1,2}, F. Damilo\textsubscript{1}, N. Forey\textsubscript{2}, G. Durand\textsubscript{2}, M.P. Vallè\textsubscript{e}\textsubscript{3}, J.L. Hopper\textsubscript{1}, M.C. Southey\textsubscript{4}, I.L. Andrilis\textsubscript{5}, E.M. John\textsubscript{6}, S.V. Tavtigian\textsubscript{2}, Breast Cancer Family Registry. 1) Genetic Cancer Susceptibility group, International Agency for Research on Cancer, Lyon, France; 2) Inserm, U900, Institut Curie, Paris, France; 3) Centre for Population Health Sciences, University of Edinburgh, UK; 4) The University of Melbourne, Victoria, Australia; 5) Institute of Medical and Environmental, Genetic and Analytical Epidemiology, School of Population Health, EGA The University of Melbourne, Victoria, Australia; 6) Cancer Prevention Institute of California, Fremont, CA, USA and Stanford University School of Medicine and Stanford Cancer Institute, Stanford, CA, USA; 7) Department of Medical Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; 8) Cancer Prevention Institute of California, Fremont, CA, USA and Stanford University School of Medicine and Stanford Cancer Institute, Stanford, CA, USA; 9) Institute of Medical and Environmental, Genetic and Analytical Epidemiology, School of Population Health, EGA The University of Melbourne, Victoria, Australia; 10) Institute of Medical and Environmental, Genetic and Analytical Epidemiology, School of Population Health, EGA The University of Melbourne, Victoria, Australia.

The RING-finger mediated \textit{BARD1}/\textit{BRCA1} heterodimer is essential for the tumor suppressor functions of \textit{BRCA1}, and both \textit{BARD1} and \textit{BARD1} may possess a pair of tandem BRCT domains that interact in a phosphorylation-dependent manner with target proteins. Since missense mutations in \textit{BRCA1} that disturb the \textit{BARD1}/\textit{BRCA1} interaction lead to breast cancer predisposition, it has been suggested that comparable mutations in \textit{BARD1} may be responsible for a proportion of familial breast cancer. Indeed, likely pathogenic germline variants in this gene have occasionally been described in breast and/or ovarian cancer families. With the aim to investigate the frequency and attributed risk of \textit{BARD1} variants in breast cancer susceptibility and to clarify relevance of mutation screening of this gene to familial cancer clinic patients, we sought to estimate the frequencies and nature of rare \textit{BARD1} variants in a sample of women with early onset breast cancer (N=1,311) and frequency matched controls (N=1,111) from three population-based centers of the Breast Cancer Family Registry. High throughput screening of \textit{BARD1} (coding and flanking intronic regions) in this series identified 30 missense variants, 13 synonymous variants and three variants in intronic regions close to splice junctions. To distinguish functionally neutral variants from those that may contribute to breast cancer risk, we used a combination of phylogenetic information with current knowledge on protein structure. Predictions on functional relevance of detected variants were obtained from different prediction algorithms. Among variants likely to alter the key RING and BRCT domains, we report for the first time a cysteine substitution (c.221G>T, p.Cys74Phe) effecting the \textit{BARD1} cancer predisposing effect. This substitution was found in one case. The only protein-truncating variant identified was the synonymous substitution c.1977 A>G (p.Arg659Arg) leading to aberrant transcript that removes the second BRCT domain of \textit{BARD1}. (p.Cys53_Trp635delinsfsX12); it was found in 7 cases (0.5%) and 9 controls (0.8%). Overall, the number of variants predicted to be damaging or neutral were evenly represented among cases and controls. Hence, our findings do not confirm in a population-based study setting of women with early onset breast cancer a risk associated with rare truncating or missense variants in \textit{BARD1}. However, we cannot exclude the possibility that mutations in \textit{BARD1} will explain some small proportion of hereditary breast cancer.

3279W

Genomic analysis of inherited breast cancer among Palestinian women. S. Lolas-Hamameh\textsubscript{1,2}, D. Dweik\textsubscript{1}, F. Fostira\textsubscript{4}, T. Walsh\textsubscript{2}, M.K. Lee\textsubscript{3}, S. Casadei\textsubscript{3}, M.C. King\textsubscript{3}, E. Levy-Lahad\textsubscript{4}, M. Kanaan\textsubscript{1}, 1) Hereditary Research Laboratory, Bethlem University, Beirut, Lebanon, Lebanon; 2) Medical Genetics institute, Sharee Zedek Medical Center, Jerusalem, Israel; 3) Departments of Medicine and Genome Sciences, University of Washington, Seattle, Washington, USA; 4) National Centre for Scientific Research "Demokritos", Athens, Greece.

In the Middle East, breast cancer incidence among Palestinian women has historically been low, but with increased education and later initiation of young child bearing among Palestinian women, is now rapidly increasing. Furthermore, perhaps because of its historically low incidence, breast cancer among Palestinian women is strikingly familial. We explored the genetic bases of this familial risk in the context of providing culturally appropriate genetic counseling services to high-risk women. Participants were 274 Palestinian and Arab-Israeli women with breast cancer, most either with a positive family history of breast or ovarian cancer (116 subjects) or diagnosed at age 40 or younger (130 subjects). Genomic DNA was tested by BRCOA, which enables the capture and simultaneous multiplex sequencing of all coding, regulatory, and intronic regions of 30 known breast and ovarian cancer genes. Considering only unambiguously damaging mutations (i.e. truncations, complete deletions, splice mutations leading to a mutant message, and missenses proven experimentally to be damaging), 29 of the 274 subjects (11%) carried a mutation responsible for their breast cancer. These included 22 of the 116 familial subjects (19%), 8 of the 130 young-onset-nonfamilial subjects (6%), and 0 of the 29 subjects not meeting either criterion. The damaging mutations included 7 in \textit{BRCA2}, 2 in \textit{ATM}, 1 in \textit{TP53}, 1 in \textit{CHEK2}, 1 in \textit{CDH1}, 1 in \textit{PALB2}, ATR, BRIP1, and \textit{XRC2}. With two exceptions (\textit{BRCA2} p.E222X and \textit{BRCA2} c.6462delTC), all mutations were different. Also, two subjects had two mutations each: in \textit{CHEK2} and \textit{BARD1}, and in \textit{CHEK2} and ATM. Multiple variants potentially altering splicing and missenses potentially damaging to function are still in process of evaluation and remain good candidates. Based on the historical demography of the region, we anticipated that the spectrum of mutations predisposing to breast cancer in the Palestinian population would be similar to the spectrum of mutations found in nonfamilial high-risk mutations, rather than a small number of founder alleles; that is, a European pattern rather than an Ashkenazi Jewish pattern. This proved to be true. Breast cancer among Palestinian women is generally diagnosed at late stages, and consequently management of early-stage breast cancer has historically been low, but with increased education and later initiation of young child bearing among Palestinian women, is now rapidly increasing.

3278F

Field synopses of genetic variation in colorectal neoplasia. J. Little\textsubscript{1}, H. Campbell\textsubscript{2}, G. Gresham\textsubscript{3}, Z. Montazeri\textsubscript{4}, S. Sivakumaran\textsubscript{5}, E. Theodoratou\textsubscript{1}. 1) Epidemiology & Community Med, Univ Ottawa, Ottawa, ON, Canada; 2) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK.

Objectives: We present results from the CRC field synopsis, report on work in progress on colorectal polyps, and comment on differences in the nature of the evidence, such as volume and quality of evidence, and issues including manner of detection of polyps, investigation of initially detected vs. recurrent polyps, and subtype. Methods: We have recently completed a field synopsis for colorectal cancer (CRC) and are in progress of developing one on colorectal polyps. For the CRC synopsis, we reviewed over 10,000 titles, then collated and extracted data from >600 publications reporting on >400 polymorphisms in >100 different genes. We carried out meta-analyses to derive summary effect estimates for >90 polymorphisms in >60 genes, including unpublished data from GWAS. For the polyp synopsis, we collated and extracted data from more than 170 publications reporting on about 200 different genes. Odds ratios are estimated based on meta-analysis for 11 polymorphisms so far. We considered four genetic models (two additives, one dominant, and one recessive). To assess the credibility of associations, we applied the Venice criteria and added consideration of Bayesian False Discovery Probability. Results: Based on meta-analysis for CRC 16 independent variants at 13 loci represent the most highly credible findings and 23 variants at 22 loci have 'less credible' association. We haven’t gene associated with polyps at this stage. In addition, we stimulate discussion about (a) updating field synopses and (b) operationalization of the Venice criteria, for both of which there appear to be differences across field synopses. These issues have relevance in public health beyond the area of genetic susceptibility. Conclusions: Our data should help direct future research effort.
Deep sequencing of prostate cancer (PrCa) genomes has recently pointed at an early role of ETS gene fusions in tumorigenesis, and revealed a characteristic landscape of structural rearrangements accompanying the oncogene translocation. Since PrCa in general has been proven partially heritable, we considered fusion positive PrCa as a distinct tumor entity that may harbor novel insights to the broadened role of MET signature genes in a range of cancers, and immediate targets for follow-up in the clinic.

### Posters: Cancer Genetics

**3280T**
Prostate cancer risk regions in 8q24 and 17q24 are differentially associated with somatic TMPRSS2:ERG fusion status. C. Maier1,2, A.E. Rickleb1,2, M. Lueddeke1,2, J.L. Stanford1,2, J. Schleskuta1,2, R.A. Eeles2,9, M. Teixeira2,10, S. Weikert11, J. Hoegel12, L.M. FitzGerald2, T. Wahlfors2, T. Visakorpi2, K.A. Leinonen12, T.L.J. Tammela13, C.S. Cooper7,14, Z. Kote-Jarai1, S. Edwards5, P. Paulo15, C. Jeronimo15,16, H. Krause1, W. Vogel1, S. Benten15, A. Alm11,16, D.F. Easton15,16, the PRACTICAL consortium.

1. Department of Urology, University of Ulm, Ulm, Germany; 2. Institute of Human Genetics, University of Ulm, Ulm, Germany; 3. Fred Hutchinson Cancer Research Center, Division of Public Health Science, Seattle, Washington, USA; 4. Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA; 5. Institute of Biomedical Technology/BioMediTech and Fimlab Laboratories, University of Tampere, Tampere, Finland; 6. Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland; 7. The Institute of Cancer Research, Sutton, UK; 8. Royal Marsden National Health Service Foundation Trust, London and Sutton, UK; 9. Department of Genetics, Portuguese Oncology Institute, Porto, Portugal; 10. Abel Salazar Biomedical Sciences Institute, Porto University, Porto, Portugal; 11. Department of Urology, University Hospital Charité, Berlin, Germany; 12. Institute of Biomedical Technology/BioMediTech, University of Tampere, Tampere, Finland; 13. Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Tampere, Finland; 14. Department of Biomedical Science, University of East Anglia, Norwich, UK; 15. Centre for Cancer Genetics Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.

**3281F**
Identification of Susceptibility Loci in Hereditary Prostate Cancer Families Using Copy Number Variation and Linkage Analysis. D. Mandal1,2, E. Leder1,2, J.E. Bailey-Wilson3,4, M.L.5.

1. Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2. Tulane Cancer Center, Tulane University School of Medicine, New Orleans, LA; 3. National Human Genome Research Institute/National Institutes of Health, Baltimore, MD; 4. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

In the United States, it is estimated that about 238,590 new cases of prostate cancer will be diagnosed and about 29,720 men will die from this disease in 2013. Family history is the most significant predictor of prostate cancer development and approximately 10% of prostate cancer cases are attributable to inheritable genetic factors. However, disease gene identification for prostate cancer has been extremely challenging due to both disease and genetic heterogeneity. To overcome the effects of ethnic disparity, genetic heterogeneity, incomplete penetrance, and missing heritability, our goal was to identify genetic components of prostate cancer by using a comprehensive approach combining both array comparative genomic hybridization (aCGH) analysis and linkage analysis. Eight prostate cancer cases were studied using aCGH to search for germ-line copy number variants (CNVs) associated with hereditary prostate cancer. The study subjects were from 8 large, high-risk, clinically homogenous families with European ancestry from Southern Louisiana. Three novel regions of CNVs were identified: 16p23, 11q22, and 2p22 in all 8 prostate cancer cases. Both model-based and model-free linkage analyses were performed on 4 of the most informative families (≥ 5 prostate cancer cases/family) from these 8 prostate cancer cases. Genotyping for linkage analyses was done using Illumina Human610-Quad HumanLinkage-12 panel. Suggestive evidence of linkage was obtained at 2q14 (HLOD score of 1.94). Using aCGH, a 68 kb duplication was observed in this region in all 8 hereditary prostate cancer cases. These genetic findings, which were identified in multiple large, well-characterized families, provide new insight into hereditary prostate cancer. Future fine mapping of the 2q14 region in a larger cohort of patients, confirmation of novel CNVs in additional subjects and use of next-generation sequencing approaches are needed to provide further evidence of susceptibility loci for prostate cancer.

**3282W**

1. Computational Medicine and Bioinformatics, Univ Michigan, Ann Arbor, MI; 2. Urology, Univ Michigan, Ann Arbor, MI.

Cancer progression is characterized, in part, by altered or abberant gene expression. Mesenchymal to Epithelial Transition (MET) is a reversible process (EMT) that is critical to metastasis in cancer progression. In this work, we established ‘MET signatures’ to characterize changes of gene expression in models of prostate cancer and breast cancer, where we induced MET by over-expression of OVOL1, OVOL2, or both OVOL transcription factors (TFs). We assessed differential expression by RNA-Seq for this set and found significant enrichment for annotation consistent with cancer progression (e.g., biological adhesion, blood vessel development). Interestingly, we also found that annotation for regulation of expression by TFs was significantly enriched in this set (FDR< 0.05 AND Fold Change ≥ 2.0) and established a set of genes representing the MET signature for each model (1,622 in BC and 2,692 in PC). The 740 genes at the intersection of these two sets represent the common MET signature for these two cancer models. We performed ConceptGen enrichment testing for this set and found significant enrichment for annotation consistent with cancer progression (e.g., biological adhesion, blood vessel development). We assessed differential expression by RNA-Seq for this set and found significant enrichment for annotation consistent with cancer progression (e.g., biological adhesion, blood vessel development). Interestingly, we also found that annotation for regulation of expression by TFs was significantly enriched in this set (FDR< 0.05 AND Fold Change ≥ 2.0) and established a set of genes representing the MET signature for each model (1,622 in BC and 2,692 in PC). The 740 genes at the intersection of these two sets represent the common MET signature for these two cancer models. We performed ConceptGen enrichment testing for this set and found significant enrichment for annotation consistent with cancer progression (e.g., biological adhesion, blood vessel development). We assessed differential expression by RNA-Seq for this set and found significant enrichment for annotation consistent with cancer progression (e.g., biological adhesion, blood vessel development). We assessed differential expression by RNA-Seq for this set and found significant enrichment for annotation consistent with cancer progression (e.g., biological adhesion, blood vessel development).
3283T

Background: Cervical carcinogenesis has two critical transition steps: persistent oncogenic human papillomavirus (HPV) infection and progression to cervical cancer. Oncogenic HPV infection alone is not sufficient to cause cervical cancer. Host genetic factors may also contribute to cervical cancer pathogenesis, but their roles remain to be determined. The defensin beta 4 gene, DEFB4, which has antimicrobial properties, is a candidate host genetic factor for susceptibility to cervical cancer. Objectives: The aim of this study was to investigate association between copy number variation of DEFB4 and susceptibility to cervical cancer in a population at high-risk of persistent oncogenic HPV infection. Methods: The study subjects comprised 204 women with cervical cancer, a population having a high-risk of persistent oncogenic HPV infection (cervical cancer group), and 200 healthy women from the general population (control group). Copy number variation of DEFB4 in each test sample was determined by relative quantitation using the comparative CT (ΔΔCT) method. Differences between the two groups were evaluated. Results: The median DEFB4 copy number in the cervical cancer group was four and in the control group was five (p=2.77e-4, 1-test). The odds ratio of cervical cancer in individuals with 4 DEFB4 copies or less was higher (odds ratio 2.02; 95% confidence interval odds ratio 1.36-3.02), compared with that in individuals with 5 or more copies (odds ratio 0.49; 95% confidence interval odds ratio 0.33-0.74). Therefore, both groups indicated a two-tailed significant difference from an odds ratio of 1.00 at the 5% level; Conclusions: We have found evidence of DEFB4 copy number variation as a host genetic factor conferring susceptibility to cervical cancer. A lower DEFB4 copy number was associated with susceptibility to cervical cancer.

3284F
Effects of Waterpipe Smoking on gene expression. Z. Montazeri, H. El katerji, J. Gomes, J. Little. Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada.

Objectives: Recently, a sharp rise in waterpipe smoking has been observed in North America and Europe, especially among young adults. There is a belief that waterpipe is the healthiest way to smoke tobacco. We studied the effects of waterpipe smoking on gene expression among young waterpipe smokers. Methods: Adverse health effects of waterpipe smoking have been reported in different studies, but this evidence has been appraised as of poor quality. There is a gap in the evidence as to the potential health effects of waterpipe smoking. We are investigating the effects of waterpipe smoking on gene expression among young waterpipe smokers in Ottawa, Canada. Results: We selected 18 genes to study the potential carcinogenic effects of waterpipe smoking on cancer based on gene expression. They have been selected based on the fact that they are induced by cigarette smoking as well as they are in the pathways of cancer diseases; genes of xenobiotic metabolism are also included. The main inclusion criteria was that the individual was between the age of 18 and 25 and reported that they smoked waterpipe. The fold change between before and after one hour and a half of smoking waterpipe, effect size, ranged between 0.02 and 34.42. Conclusions: Results could be used to predict the health effects of waterpipe smoking. This research will be used in knowledge translation to enable public health professionals and policy makers to make informed decisions about the control of waterpipe smoking, including potential prevention strategies and cessation interventions.

3285W
Upregulation of TRF1 and TRF2 (telomere repeat binding factors) protein contributes to telomere shortening in renal cell carcinoma. D. Pal1*, U. Sharma1, R. Khajuria3, S.K. Singh2, N. Kakkar2, R. Prasad1. 1) Dept OB/GYN, PGIMER, Chandigarh, Chandigarh, India; 2) Urology, PGIMER, Chandigarh, India; 3) Histopathology, PGIMER, Chandigarh, India.

Upregulation of TRF1 and TRF2 (telomere repeat binding factors) protein contributes to telomere shortening in renal cell carcinoma (RCC). Materials and Methods: Total 80 cases of RCC treated by surgery under advanced Urology services of Nehru Hospital, at Postgraduate Institute of Medical Education and Research, Chandigarh were included in the present study. For comparison, normal renal cortex samples were taken in each case. Transcriptional expression of TRF1 and TRF2 were estimated by real time PCR. Whereas Protein levels were detected by using immunohistochemical and immunofluorescence method. The mean telomere length was determined by southern blotting followed by hybridization. Results: The expression of TRF1 and TRF2 were significantly higher in RCC tissue in comparison with normal renal parenchyma. The mean telomere length in RCC tissue was significantly shorter than that in normal renal parenchyma. The mean telomere length in all tissue samples were inversely correlated with the level of TRF1 and TRF2 expression. Conclusions: Our result suggests that the upregulation of TRF1 and TRF2 may work to reduce the telomere length in RCC and could contribute to the carcinogenesis of renal cell carcinoma.

3286T
Screening of XRCC2 in breast cancer families. L.M. Peilttari1, S. Vitsko1, C. Blomqvist2, K. Aittomäki3, H. Nevanlinna1. 1) Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 2) Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 3) Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland.

XRCC2, one of the five RAD51 paralogs, is involved in the repair of DNA double-strand breaks through homologous recombination. Most of the breast cancer susceptibility genes, including the high risk genes BRCA1 and BRCA2, have an important role in DNA damage repair. Deleterious mutations in the XRCC2 gene were recently identified in an exome-sequencing study of familial breast cancer patients and a homoygous truncating mutation in the gene was identified in a Fanconi anemia patient. However, the association of XRCC2 with breast cancer was not confirmed in a larger case-control study. To investigate the role of XRCC2 mutations in breast cancer predisposition in the Finnish population, we sequenced the coding region, the 5'UTR, and the exon-intron boundaries of the gene in blood DNA samples of 344 familial breast cancer patients. This extensive set of non-BRCA1/2 breast cancer families included 342 families with at least three breast or ovarian cancer patients among first or second degree relatives and 2 families with two affected first degree relatives. We detected four known polymorphisms among the breast cancer families: one downstream, one 5'UTR, one intronic and one missense variant. No novel changes or truncating mutations were identified. None of the polymorphisms were predicted to be pathogenic and the minor allele frequencies were comparable to the population frequencies. Our results indicate that XRCC2 does not contribute to familial breast cancer predisposition in the Finnish population. Taken together, it is unlikely that XRCC2 has a significant contribution to breast cancer susceptibility.
3287F
Large-scale resequencing analysis of six melanoma susceptibility genes in the European Prospective Investigation into cancer and nutrition Cohort. M. Pertesi1, N. Forey1, J. Oliver1, N. Robint1, C. Voegele1, EPIC, skin cancer working group2, F. Le Calvez-Kelm1, J. McKay1, F. Lesueur1,2. 1) Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, Lyon, France; 2) Inserm U900, Institut Curie, Mines ParisTech, Paris, France; 3) International Agency for Research on Cancer, Lyon, France.

Cutaneous malignant melanoma (CMM) is an important health problem in fair-skinned populations worldwide, showing a dramatic increase in incidence over the past decades. Established risk factors include sun exposure, nevus propensity, pigmented traits, and familial history. Inherited DNA sequence variants are involved in the development and progression of CMM and three classes of susceptibility genes have been described. Highly penetrant mutations in the cell cycle related 1 genes CDKN2A and CDKN2B account for about 2% of CCM cases across populations, while more moderately penetrant alleles (MC1R, MITF), and very low penetrance SNPs in genes related to pigmentation, nevus count, immune response, DNA repair, and metabolism explain about 12% of the familial risk in European populations. Given that the genome-wide association studies conducted by International consortia were well powered to detect SNPs with OR>1.5, it seems unlikely that many additional common alleles with moderate effects on melanoma risk exist, indicating that other uncommon-to-rare susceptibility alleles remain to be identified. To test this hypothesis we designed a cost-effective multiplex-targeted resequencing assay specifically testing for 6 melanoma susceptibility genes identified in the pangenomic studies. About 17Kb of DNA sequence corresponding to the exons and flanking intronic regions of MC1R, MITF, ASIP, TYR, TPR1 and BAP1 genes is being screened in 1061 CMM cases, and 1061 matched controls from the European Prospective Investigation into cancer and nutrition Cohort. Our custom mutation screening workflow involves the creation of a barcoded sample library for each individual including amplified regions of interest and sequencing on the PGM Ion Torrent sequencer. Each step of the process is tracked by a Laboratory Information Management System. To assess the sensitivity and specificity of our custom workflow, we compared PGM data to the HapMap and Sanger data in a subset of 429 subjects screened for variants in the highly polymorphic MC1R gene. We identified 462 and 461 sequence variations, with the two workflows, respectively. These correspond to 27 distinct common and rare variants. These preliminary data suggest that the resequencing approach, as applied in this study, is robust and capable of uncovering a substantial number of genetic variants for melanoma.

3287T
COMPLEXO: Identifying the missing heritability of breast cancer via next generation collaboration. M. C. Southey1, D. E. Goldgar2, COMPLEXO, Melanoma and Skin Cancer Research Group, Melbourne, Melbourne, VICTORIA, Australia; 2) Huntsman Cancer Institute, The University of Utah School of Medicine, Salt Lake City, UT 84112, USA.

A proportion of the remaining unexplained genetic susceptibility to breast cancer is likely to reside in rare variants of many small effect size genes. Coordinated international collaboration offers great potential to advance the discovery of additional breast cancer susceptibility genes by increasing the likelihood of identifying functionally relevant genetic variants in rare variants in multiple families. A new consortium, COMPLEXO (a name chosen to reflect the complexity of the exome), has been formed to facilitate collaborations between researchers actively applying massively parallel sequencing to understand the genetics of breast and ovarian cancer. The aim of COMPLEXO is to bring to massively parallel sequencing the same power of large sample sets that have proven so successful in examining the role of common variants in cancer populations via the consortium model such as the Breast Cancer Association Consortium (BCAC), the Ovarian Cancer Association Consortium (OCAC) and the Collaborative Oncology Lung Disease (COLAD) consortium. In the EPIC Gene-environment Study (COGES), sequencing studies provide additional challenges in terms of defining specific modes of collaboration differ in sequencing and targeted capture platforms, bioinformatics platforms, the need to integrate ongoing studies in many centers and the socio-ethical-legal issues that are not as relevant to initiatives that are genotyping common genetic variation. These issues are relevant to research in all complex human diseases. We will describe our working and governance structures, our early experiences and ongoing activities aimed at identifying more of the missing heritability of breast cancer.

3288W
K939Q polymorphism in the XPC gene may not affect the risk of bladder cancer: a case-control study and a meta-analysis. M. SANKHWAR1, S.N. SANKHWAR1, N. GUPTA1, A. ABHISHEK1, S. RAJENDER1. 1) Urology, King George Medical University, Lucknow, Uttar Pradesh, India; 2) Endocrinology Division, CSIR-CDRi, Lucknow, Uttar Pradesh, India.

Purpose: Lys939Gln (K939Q), an A>C transversion in exon 15 of the XPC gene, has been extensively studied in relation to cancer. Keeping in view its decisive polymorphic effects and colocations with bladder cancer, we have analysed this polymorphism in Indian bladder cancer patients, and its decisive polymorphic effects and collocations with bladder cancer, we have analysed this polymorphism in Indian bladder cancer patients, and the prothrombin (rs1799963) variants. Genotype-phenotype correlations were conducted between the SNPs that were associated with breast cancer and mRNA expression in tumor and plasma levels of twelve hemostatic parameters. Increased levels of the coagulation factors F8, F9 and VWF, and decreased levels of antithrombin were detected in the patients. Global activation of coagulation was demonstrated in an endogenous thrombin generation assay, by increased D-dimer levels, and increased activated protein C resistance (APCR). The genetic risk factors of thrombosis (F5, F2, F3, F5, F7, F10 and TFPI) and the prothrombin (rs1799963) variants, showed no association with breast cancer. Four common SNPs in F5 and one in F10 were significantly associated with risk of breast cancer. The minor alleles of F5 rs12120605 and rs6427202 were more frequent in patients (OR=1.52, [95% CI=1.07-2.17] and OR=1.40, [95% CI=1.02-1.91]). They were associated with increased tissue factor factor and factor 5 mRNA expression in tumors, respectively. In contrast, the minor alleles of F5 rs6427199 and rs9332542 were more frequent in controls (OR=0.71, [95% CI=0.52-0.96] and OR=0.65, [95% CI=0.49-0.87]), and they correlated with antithrombin in plasma and factor V mRNA levels in tumor, respectively. This implies that these SNPs may be located in trans- and cis-acting DNA elements that contribute to regulate the pro-coagulant expression in plasma and tumors. The F5 rs6427199 also correlated with lowered APCR, partly explaining the case-control association by the F5 rs9332542. F10 rs3093261 also associated with clear effect on breast cancer (OR=1.60, [95% CI, 1.17-2.17]), but with no clear effect on hemostatic phenotypes. In conclusion, this study demonstrates novel associations of common SNPs in hemostatic genes (F5 and F10) influencing the pro-coagulant phenotype verified in patients may be due to genetic effects.

3289F
Common genetic polymorphisms in the coagulation factors 5 and 10 genes are associated with risk of breast cancer and correlate with increased coagulation activity. M. Tinholle1,2,1, M. Viken1, S. Nyberg2, AL. Barrenes-Dalé1,4, V. Kristensen4,5,6, KK. Sahliberg4,5,6, R. Kåresen1,7, E. Schlichting1, O. Garred8, G. Skretting2, AE. Dahm9, BA. Lie1, PM. Sandset9, N. Iversen1. 1) Dept. of Medical Genetics, Oslo University hospital and University of Oslo; 2) Dept. of Haematology and Research Institute of Internal Medicine, Oslo University Hospital; 3) Institute of Clinical Medicine, University of Oslo; 4) Dept. of Genetics, Institute for Cancer Research, Oslo University Hospital; 5) The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo; 6) Dept. of Clinical Molecular Biology (EpiGen), Akershus University Hospital, Oslo; 7) Dept. of Breast and Endocrine Surgery, Oslo University Hospital; 8) Dept. of Pathology, Oslo University Hospital.

Tumor cells express coagulation factors, thereby creating a pro-thrombotic state in cancer. Besides increasing the risk of thrombosis, pro-coagulants may promote pro-cancer signaling independent of coagulation. We aimed to gain a better understanding of the molecular mechanisms controlling the interaction between breast cancer and coagulation. We performed a study of 390 non-treated breast cancer patients and 350 healthy controls. SNP associations in the hemostatic genes F2, F3, F5, F7, F10 and TFPI (n=42 from HapMap) were explored, in addition to the F5 Leiden (rs60252) and the prothrombin (rs1799963) variants. Genotype-phenotype correlations were conducted between the SNPs that were associated with breast cancer and mRNA expression in tumor and plasma levels of twelve hemostatic parameters. Increased levels of the coagulation factors F8, F9 and VWF, and decreased levels of antithrombin were detected in the patients. Global activation of coagulation was demonstrated in an endogenous thrombin generation assay, by increased D-dimer levels, and increased activated protein C resistance (APCR). The genetic risk factors of thrombosis (F5, F2, F3, F5, F7, F10 and the prothrombin variants) showed no association with breast cancer. Four common SNPs in F5 and one in F10 were significantly associated with risk of breast cancer. The minor alleles of F5 rs12120605 and rs6427202 were more frequent in patients (OR=1.52, [95% CI=1.07-2.17] and OR=1.40, [95% CI=1.02-1.91]). They were associated with increased tissue factor factor and factor 5 mRNA expression in tumors, respectively. In contrast, the minor alleles of F5 rs6427199 and rs9332542 were more frequent in controls (OR=0.71, [95% CI=0.52-0.96] and OR=0.65, [95% CI=0.49-0.87]), and they correlated with antithrombin in plasma and factor V mRNA levels in tumor, respectively. This implies that these SNPs may be located in trans- and cis-acting DNA elements that contribute to regulate the pro-coagulant expression in plasma and tumors. The F5 rs6427199 also correlated with lowered APCR, partly explaining the case-control association by the F5 rs9332542. F10 rs3093261 also associated with clear effect on breast cancer (OR=1.60, [95% CI, 1.17-2.17]), but with no clear effect on hemostatic phenotypes. In conclusion, this study demonstrates novel associations of common SNPs in hemostatic genes (F5 and F10) influencing the pro-coagulant phenotype verified in patients may be due to genetic effects.
3291W
Estrogen and Progesterone receptor gene polymorphisms as risk factors for Cervix Cancer in South Indian Population. P. Upendram1, V. Kiran Kumar1, S. Poornima1, A. Shah1, Q. Hasan1. 1) Department of Genetics & Molecular Biology, Kamineni Hospitals, Hyderabad, India; 2) Department of Oncology, Kamineni Hospitals, Hyderabad, India.
Carcinoma of the uterine cervix is the most frequent gynaecological malignancy affecting women in developing countries despite being a potentially preventable disease. Globally, it is the second most frequent cancer affecting women with a steady rise in incidence among younger women. It is evident from current literature that specific polymorphisms in hormone receptor genes may confer an increased risk dependent on certain hormones. This study was designed to evaluate the responses to estrogen and progesterone lead to steroid hormone dependent disorders like endometriosis, endometrial cancer, breast cancer and ovarian cancer. Despite evidence that estrogen may play an important role in the carcinogenesis of cervical cancer, its action and mechanism in cervical cancer invasion is not well defined. To the best of our knowledge there are no studies that have focused on the estrogen and progesterone receptor (ER and PR) gene variants in cervical cancer especially in the Asian Indian population. The aim of this case control study from South India was to establish the association between the T/C SNP of estrogen receptor (ER) alpha gene recognized by PvuII enzyme (rs2234693) and a 306bp Alu insertion in the PGR gene (rs1042838) with cervical cancer. DNA was isolated from a total of 280 women including 100 cases with cervical cancer and 100 age-matched healthy controls using the salting out technique. Polymorphisms were evaluated by Polymerase Chain Reaction followed by restriction enzyme digestion / gel electrophoresis. Our results showed a statistically significant association of the C allele of ER gene (OR = 5.8072; 95% CI: 1.3700-23.13; p=0.0001) and the T allele of PGR gene (OR = 2.5362; 95% CI: 1.0800-5.9533; p=0.0312) with cervical cancer. These findings are implicated in the promotion and prevention of various cancers. The uterine cervix is highly responsive to estrogen, however, the role of estrogen receptor polymorphisms in cervical cancer has not been well established. Our results suggest that the ER/PR polymorphism but not PROGINS polymorphism can be used as a bio-marker to identify women with a high risk of developing cervical cancer. A larger study in different ethnic groups is warranted for establishing the association of cervix cancer with the ER polymorphism.

3292T
Fine-mapping of Genome-wide Association Study-identified Risk Loci for Colorectal Cancer in African Americans. H. Wang1, C.A. Haiman2, T. Burnett1, B.K. Fortini3, L.K. Kolonel1, B.E. Henderson1, L.B. Sigmond4,5, W.J. Blot3,4, T.O. Keku5, S.I. Bembout4, P.A. Newcomb3, M. Pande3, D.O. Stram1, L. Le Marchand1. 1) University of Hawaii Cancer Center, Honolulu, HI; 2) Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 3) International Epidemiology Institute, Rockville, MD; 4) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN; 5) Center for Comprehensive Cancer Biology of North Carolina, Chapel Hill, NC; 6) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 7) Fred Hutchinson Cancer Research Center, Seattle, WA; 8) Department of Epidemiology, the University of Texas M. D. Anderson Cancer Center, Houston, TX; 9) Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, NH; 10) Cancer Prevention Institute of California, Fremont, CA.

Genome-wide association studies of colorectal cancer (CRC) in Europeans and Asians have identified 21 risk susceptibility regions (29 index SNPs). Characterizing these risk regions in diverse racial groups with different LD structure can help localize causal variants. We examined associations between CRC and all 29 index SNPs in 6,597 African Americans (1,894 cases and 4,703 controls). Nine SNPs in 8 regions (5q31.1, 6q26-q27, 8q23.3, 8q24.21, 11q13.4, 15q13.3, 18q21.1, 20p12.3) formally replicated between CRC and all 29 index SNPs in 6,597 African Americans (1,894 cases and 4,703 controls). Among the SNPs correlated with the index variants, two markers, rs12759486 (or rs7544751), a putative functional SNP in and near HOXB13, and rs1801131, were more strongly and statistically significantly associated with CRC, after conditioning on the index signals (p < 0.0006). The average per allele risk was improved using the replicated index variants and the two new markers (OR = 1.14, 95% CI: 1.10-1.18; p = 3.4 x 10−16). The contribution of the two new risk SNPs to CRC heritability was estimated to be 1.5%. This study highlights the importance of fine-mapping in diverse populations.

3293F

Introduction: Laboratory studies have demonstrated that zinc and zinc transporters play important roles in the development of urological cancers, including prostate, bladder, and renal cell cancer. To date however, there have been no efforts to evaluate whether germline variation in zinc transporter genes alter the risk of developing these cancers. Motivated by this, we evaluated associations of known variants in zinc transporter genes with risk of prostate, bladder, and renal cell cancer using three NCI Genome Wide Association Study (GWAS) data sets downloaded from dbGaP. Methods: Data consisted of GWAS data from the PLCO Screening Trial (prostate; 1172 cases, 1157 controls); Bladder Cancer study (3527 cases, 5119 controls); and Renal Cancer study (1453 cases, 3531 controls). All genotyped variants in 10 ZNT family genes (ZNT1-ZNT10) and 14 ZIP family genes (ZIP11-ZIP14) in each data set were analyzed by logistic regression models to evaluate the association of each variant with risk of developing prostate, bladder, and renal cancer, respectively. We evaluated these associations in a univariable setting, after adjusting for available covariates, such as age and gender. Correction for multiple comparisons used Max T permutation tests. Results: One variant in ZIP11, rs8081059, was associated with increased risk of renal cancer (OR=1.28, 95% CI: 1.13-1.45; corrected p=0.049). Another ZIP11 variant, rs1871756, was associated with an increased risk of bladder cancer (OR=1.43, 95% CI: 1.24-1.63; corrected p=0.0002). We also found three variants in ZIP11 associated with a decreased risk of bladder cancer: rs11077654 (OR=0.76, 95% CI: 0.68-0.85; corrected p=0.001), rs9913017 (OR=0.76, 95% CI: 0.68-0.85; corrected p=0.002), and rs4389054 (OR=0.78, 95% CI: 0.63-0.98; corrected p=0.02). Interestingly, these three protective variants are within the same LD block. There were no associations between zinc transporter variants and prostate cancer risk. Conclusion: We report for the first time that specific genetic variants of a zinc transporter gene, ZIP11, are associated with risk of renal and bladder cancer. If validated in larger, more focused studies, further investigations are warranted to explore the functional significance of these findings in ZIP11 in these two cancers.

3294W
The Prevalence of the HOXB13 G84E Prostate Cancer Risk Allele in Men Treated with Radical Prostatectomy, K.A. Zuhlike1, J. Beebe-Dimmer2, C. Yee2, A.M. Johnson1, P.C. Walsh3,4, S.D. Isaacs5,6, W.B. Isaacs5,6, K.A. Cooney1,2. 1) University of Michigan Medical School, Department of Internal Medicine, Ann Arbor, MI; 2) Wayne State University, Department of Oncology, Detroit, MI; 3) Karmanos Cancer Institute, Detroit, MI; 4) Johns Hopkins University, Baltimore, MA; 5) James Brady Urologic Institute, Baltimore, MD; 6) University of Michigan Medical School, Department of Urology, Ann Arbor, MI.

Purpose: A rare, non-conservative substitution (G84E) in the homeobox transcription factor HOXB13 gene, located on chromosome 17q21, was recently reported to confer an increased risk of prostate cancer. As a follow-up to this investigation, DNA samples from 9,559 men undergoing radical prostatectomy were genotyped to determine the frequency and clinical correlates of this mutation in men with prostate cancer.

Materials and Methods: DNA samples from men treated with radical prostatectomy at the University of Michigan and John Hopkins University were genotyped for G84E and confirmed by Sanger sequencing. The frequency and distribution of this allele was determined according to specific patient characteristics (family history, age at diagnosis, pathologic Gleason grade and stage).

Results: 128 of 9,559 patients were heterozygous carriers of G84E (1.3%). Patients who possessed the variant were more likely to have a family history of prostate cancer (46.0% vs. 35.4% p=0.006). G84E carriers were more likely diagnosed at a younger age compared to non-carriers (55.2 years vs. 58.1 years; p=0.0001). However, there was no difference in the proportion of patients diagnosed with high-grade or advanced stage tumors between carrier status.

Conclusions: In our study, carriers of the rare G84E variant in HOXB13 were both younger at the time of diagnosis and more likely to have a family history of prostate cancer compared to homozygotes for the wild-type allele. No significant differences in allele frequency were detected according to race or ethnicity. Further investigation is required to evaluate the role of HOXB13 in prostate carcinogenesis.
3295T
Prevalence of germline mutation p.R337H in the TP53 gene in families with multiple cases of cancer. K.C. Andrade1, 2, K.M. Santiago1, A. Nobrega1, F.P. Fortes1, M.I. Achata1, A.C. Camargo Cancer Center, São Paulo, Brazil; 2) UNESP, São Paulo State University - Biosciences Institute, Botucatu - São Paulo - Brazil.

BACKGROUND: Germline mutations in the TP53 gene are associated to the Li-Fraumeni Syndrome (LFS) and its variants Li-Fraumeni-Like (LFL). The diseases predispose carriers for early onset tumors, including soft tissue sarcoma, pre-menopausal breast cancer, central nervous system tumors and adrenocortical carcinoma. In Brazil, there is a high frequency of a germline TP53 mutation in South and Southeastern populations due to a founder effect. It is estimated to be present in 0.3% of the local inhabitants. However, only a few families have been diagnosed with such alteration. Our hypothesis was that the current criteria applied for LFS are not enough to detect them. Therefore, the development of suitable criteria to select them will enable not only the identification of at-risk families, but also provide adequate screening and early detection of associated tumors. OBJECTIVE: To determine the efficacy of wide criteria for detection of p.R337H carriers.

METHODS: 31 patients from Oncogenetics Department, A.C. Camargo Cancer Center, SP, were selected and tested for the p.R337H mutation. Criteria for inclusion were: (1) more than three family members with cancer AND (2) at least one of them under age 50 AND (3) two of them being first or second degree relatives. Germline DNA was extracted from peripheral blood and analyzed by Restriction Fragment Length Polymorphism. The confirmation of positive finding was done through direct sequencing of exon 10.

RESULTS: One out of 31 patients (3.22%) was found to carry the p.R337H mutation. The patient developed ductal invasive breast cancer at age 47 and invasive adenocarcinoma of the lung at age 48. In addition, an extensive cancer family history was reported. DISCUSSION: These results show that the proposed criteria may detect carriers, but further studies including a larger group of families will be useful to define its effectiveness. Also, in order to improve sensitivity, the inclusion of more stringent criteria will be needed. Moreover, one of the tumors manifested in family history is of great interest: the development of an Ewing's Sarcoma (ES) in her second degree relative at age 9. Current data indicate that the annual incidence of ES has hiked in the last decade of life and lesions in the GH-producing tumor from the patient was also detected in healthy controls. The increased expression of IGSF1 was quite polymorphic. The mutation identified in one of our patients with giantism was also detected in healthy controls. The increased expression of IGSF1 showed variable IGSF1 staining in the GH-producing tumor from the patient and invasive adenocarcinoma of the lung. The same variation, however, was found in drug and steroid hormone metabolism, and PARP2 is known to cause Bloom Syndrome, a recessive disorder with multiple phenotypic effects including predisposition to various cancers. Additionally, a BLM truncating mutation was recently identified as a possible tumor-suppressor gene and a PARP2 polymorphism has been associated to breast cancer. This suggests that IGSF1 may be a modifier during oncogenesis.

3296F
Is IGSF1 involved in human pituitary tumor formation? F.R. Fauz1, A.D. Horvath2, P. Xekouki1, E. Szarek1, G. Evgenia1, A.D. Manning1, I. Levy, E. Saloustros3, R.B. de Alexandre1, M. Nesteiro1, C.A. Stratakis1, 1) Section on Endocrinology & Genetics, Program on Developmental Endocrinology & Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD; 2) McCombie Genomic and Proteomic Center, Department of Biochemistry and Molecular Biology, The George Washington University, Washington, DC.

IGSF1 is a membrane glycoprotein highly expressed in the anterior pituitary gland. Mice with a deficiency in Igsf1 show central hypothyroidism and increased body size. Recently, mutations in this gene were found in humans with central hypothyroidism and testicular enlargement [Sun et al. Nat. Genet. 44(12):1375-81, 2012]. We looked for IGSF1 germline variations in patients with gigantism from the NIH data registry and in healthy controls. Germline DNA was extracted from peripheral blood and analyzed by Restriction Fragment Length Polymorphism. The confirmation of positive finding was done through direct sequencing of exon 10. Only one of the 11 patients (100% White Americans) for IGSF1 germline mutations. Immunohistochemistry for IGSF1 was performed in sections from three GH-producing adenomas and in normal pituitary. In 1 out of 5 patients, we identified the sequence variant p.Asn604Tr (c.1811A>C), which by in silico analysis is potentially deleterious. The same variation, however, was found in 2 of our controls. We also identified 29 more variations in a total of 18 healthy individuals (16.6%) (3 nonsense, 3 frameshifts and 25 missense), but only 4 of them have been described previously. Immunohistochemistry showed variant IGSF1 staining in the GH-producing tumor from the patient with the IGSF1 mutation compared to a GH-producing tumor from a patient negative for IGSF1 mutations and to a normal control. IGSF1 gene appears quite polymorphic. In an attempt to explain this variation, one of the authors was also detected in healthy controls. The increased expression of IGSF1 in patients with an IGSF1 germline mutation may indicate that IGSF1 does not have a causative role in pituitary tumor development but may work as a modifier during oncogenesis.

3297W
Mutational Landscape of Candidate Genes in Familial Prostate Cancer. A. Johnson1, K. Zhulke1, C. Plotts2, J. Douglas3, S. Thibodeau4, K. Cooney3, 1) Internal Medicine - Hem/Onc, Univ Michigan, Ann Arbor, MI; 2) Human Genetics, Univ Michigan, Ann Arbor, MI; 3) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 4) Urology, Univ Michigan, Ann Arbor, MI.

Family history is a major risk factor for prostate cancer (PCA), suggesting a genetic component to this disease. However, traditional linkage and association studies have failed to fully elucidate the underlying genetic basis of familial PCAs. Here we use a candidate gene approach to identify potential PCAs susceptibility loci in whole-exome sequencing data from familial PCAs cases. We identified 94 candidate genes in 31 multiplex PCA families. All 351 candidate genes were shared by all 3 families in each family in which they were observed; 43 of these were observed in all 33 cases and were presumed to be rare variants in the reference sequence and excluded from further analysis. The remaining 13 shared, missense variants as well as the 9 nonsense variants were re-sequenced for verification. Confirmed variants were then tested in all additional family members with available DNA to assess co-segregation with PCA status. All 9 nonsense variants and 12 of 13 shared, missense variants were confirmed. Of the 9 nonsense variants was HOXB13 Gly44Glu which was recently identified as a PCA susceptibility allele. The Gly44Glu minor allele was observed in 1 family where it showed partial co-segregation with disease. Two other missense variants, BLM Gin123Arg and PARP2 Arg283Gln, and 1 nonsense variant, CYP1A1 Asn138Ser, was observed in at least one case. While previous data suggested that a NGS analysis is potentially destructive. POSTERS: Cancer Genetics

3298T
Development of a Next Generation Sequencing panel to assess hereditary cancer risk that includes clinical diagnostic analysis of the BRCA1 and BRCA2 genes. B. Roa1, K. Bowles1, S. Bhatnagar1, N. Gutin1, A. Murray1, B. Wardle2, M. Bastian1, J. Mitchell1, J. Chen2, T. Tran2, D. Williams2, J. Potter2, S. Jammulapati2, M. Perry2, B. Morris2, K. Timms2, 1) Myriad Genetic Laboratories, Inc., Salt Lake City, UT; 2) Myriad Genetics, Inc., Salt Lake City, UT.

Approximately 7% of breast and 11-15% of ovarian cancers are estimated to be due to germline DNA mutations, the majority of which occur in the BRCA1 and BRCA2 genes. However, a comprehensive hereditary breast and ovarian cancer risk assessment should include germline sequencing of additional genes associated with breast cancer risk, including those with a known association in breast/ovarian cancer patients. Sanger DNA sequencing has been the gold standard for molecular genetic analysis. However, it is labor-intensive and costly for the analysis of large gene panels. Next Generation Sequencing (NGS) platforms allow for efficient analysis of larger gene panels, but lack of standardization of sample preparation, NGS platforms, and data analysis presents challenges to diagnostic laboratories. Optimized assay design and validation are critical to maximize the analytical sensitivity of NGSS assays, and to ensure high quality interpretation for clinical decision making. We developed a 25-gene NGS hereditary cancer panel that uses RainDance PCR technology for high-throughput sample preparation, Illumina HiSeq and MiSeq NGS technologies, and commercially available and lab-developed informatic tools. The panel and assays were custom designed to ensure high-quality, clinically actionable results. Initial assessment of analytical sensitivity and specificity was performed by comparing BRCA1 and BRCA2. NGS was performed on 1864 anonymized patient samples, which had previously undergone Sanger sequencing of BRCA1 and BRCA2. Sensitivity of which 13 mutations, 24 SNPs and 93 SNVs were unique and 482 were classified as disease-associated mutations. We identified 15,877 variants at an initial sensitivity of >99.99% for BRCA1 and BRCA2. One polymorphic variant was missed due to a variant under the threshold. The NGS assay was successfully optimized through further improvements. No additional variants were found by NGS, yielding a specificity of 100%. This preliminary analysis facilitated assay optimization, and a comprehensive validation of all 25 genes in the NGS panel. This analysis indicates that a NGS gene panel can provide high quality interpretation for clinical decision making.
Development of 3C-based target sequencing technology for candidate gene discovery in prostate cancer. MJ. Du1, TZ. Yuan1, RL. Dittmar1, XY. Huang1, SN. Thibodeaux2, L. Wang1. 1) Pathology, Medical College of Wisconsin, Milwaukee, WI; 2) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

So far, genomewide association studies have identified over 70 chromosome loci (represented by single nucleotide polymorphisms, SNPs) with an increased risk of prostate cancer. However, the biological roles of these loci in prostate cancer development are unclear. As many of these risk-loci have been found in non-coding regions of the genome, with many residing at some distance from any nearby annotated genes, many of these causal variants are believed to regulate target genes through long range interactions. Although chromosome conformation capture (3C) is used to analyze the long range interactions between different fragments, the traditional 3C assay and its variants are not sufficient to identify unknown genomic regions that interact with multiple risk loci in a cost-efficient manner and in a single enrichment assay. Here, we developed a 3C-based multiple targets enrichment assay and applied this assay to the analysis of 3 prostate cancer risk regions at 8q24. The novelty of this assay is to use short oligos as primers and label target fragments by multiple primers extension. Briefly, a standard approach was used to generate a 3C library by using LinCap cell line, followed by sonication and size selection of 3C DNA into 300-500bp fragments. Nine short oligos (in three risk regions at 8q24) were then hybridized to the fragmented 3C library. Primers extension at high annealing temperature (65°C) was performed with the presence of biotin-labeled dCTP. After strepavidin beads purification, a nano scale library construction method was applied to build a sequencing library. qPCR analysis showed up to 10 fold target enrichment under our current condition. Finally, a perl based computer algorithm was written to analyze the long range interactions from large scale sequencing data. The sequencing data showed that the risk regions interacted with multiple chromosome loci. However, interactions with MYC gene, 250kb downstream from nearest prostate cancer risk SNP, were most prominent. Nine short oligos (in three risk regions at 8q24) were then hybridized to the fragmented 3C library. Primers extension at high annealing temperature (65°C) was performed with the presence of biotin-labeled dCTP. After strepavidin beads purification, a nano scale library construction method was applied to build a sequencing library. qPCR analysis showed up to 10 fold target enrichment under our current condition. Finally, a perl based computer algorithm was written to analyze the long range interactions from large scale sequencing data. The sequencing data showed that the risk regions interacted with multiple chromosome loci. However, interactions with MYC gene, 250kb downstream from nearest prostate cancer risk SNP, were most prominent.

1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Cancer and Cell Biology Division, The Translational Genomics Research Institute (TGen), Scottsdale.

There are multiple promising new immunotherapeutic and targeted approaches for treating melanoma, not all patients are eligible for these therapies and most of those that are ultimately relapse, highlighting a critical need for improved chemotherapeutics. Temozolomide (TMZ) is an orally administrating agent used in the treatment of advanced cancers including melanoma. TMZ has shown some advantage in the clinic in terms of preventing central nervous system relapse and patient quality of life. In order to find novel vulnerabilities that may be exploited to sensitize melanoma cells to TMZ, we utilized pre-existing expression profiling data and known TMZ-sensitivity from 18 melanoma cell lines to develop a classifier of TMZ-sensitivity. We used the support vector machine (SVM) method to build the model, and applied this classifier to predict the TMZ-sensitivity in 25 independent melanoma cell lines. We subsequently validated these predictions by experimentally determining the IC50 of these cell lines to TMZ using a high-throughput cell viability assay. The result showed that more than 80% of melanoma cell lines were accurately predicted. We analyzed the genes differentially up-regulated by TMZ sensitive and -resistant melanoma cells using both microarray data, and most recently RNA-sequencing data using Ingenuity Pathways Analysis. We identified two key genes (MTIF and EGFR) each known to regulate a large number of the differentially expressed genes. EGFR and a large network of EGFR targets were significantly underexpressed in TMZ resistant lines. At the same time, MTIF and a large group of known MTIF targets were significantly overexpressed in resistant lines. We also identified E-cadherin (CDH1), which is known to play a key role in invasion and transition from radial to vertical growth in melanoma, as highly upregulated in resistant cell lines. We functionally characterized the role of EGFR in TMZ sensitivity via EGF treatment or overexpression in resistant cell lines, as well as antibody, small molecule, and shRNA inhibition of EGFR in sensitive cell lines. Small-molecule inhibition of EGFR decreases downstream ERK and AKT pathway signaling in both TMZ-resistant and TMZ-sensitive cell lines tested. Our expression profiling data suggest that TMZ sensitivity may be mediated by multiple large gene regulatory networks, and that EGFR may play a role in therapeutic response of melanoma to TMZ.


NextBio, Santa Clara, CA.

We previously identified AGR3, a protein disulfide isomerase with unknown physiological function, as an RNA expression marker for triple negative breast cancer based on ESR1/PGR/ERBB2 RNA levels using NextBio (NB) Clinical, a patient-centric analytical platform. Here we investigated AGR3’s role in 19 cancers by mining genomic data of 11,000 patients currently integrated into NB Clinical and performed correlations against publically available curated genomic data using NB Research. Analysis of RNA expression data from patients of 19 cancers revealed AGR3 dysregulation in ovarian, kidney, lung, head/neck, breast, uterine, and brain cancers, using a 2-fold cutoff in either direction as compared to normal tissue. These patients were further stratified based on AGR3 over vs underexpression, using a 2-fold cutoff against disease median. Fifty top-ranked dysregulated genes from each cancer totaling 291 genes were identified and imported to NB Research for further insights. The NB Body Atlas application showed our 291 gene signature to most highly correlate with lung amongst all normal tissues, and highest with bronchial and airway epithelial cells amongst cell types. AGR3 itself also had the highest correlation with pulmonary epithelial cells, and was highly correlated with lung cancer when queried with the NB Disease Atlas application. We followed up by assessing the top 50 genes obtained with NB Clinical from lung cancer patients alone. Stratification by AGR3 expression showed co-amplified expression of genes involved in pulmonary surfactant homeostasis and secretion, including surfactant proteins (SFTPA-D), surfactant associated-proteins (SFTA1-2), and secretoglobin (SCGB1A1, SCGB3A1-2). Analysis of patient clinical parameters showed decreased AGR3 expression in 66% of current smokers and in 71% of patients with the primitive subtype of lung squamous cell carcinoma (LSCC). AGR3 was increased in 65% of patients with secondary LSCC. Through analysis of individual patients and correlations with public genomic data, we identify dysregulated pulmonary epithelial AGR3 expression as a biomarker for cancer with an effect on surfactant homeostasis. This is functionally akin to its paralog, AGR2, which partakes in intestinal mucus production and is reported to be overexpressed and secreted in prostate and pancreatic cancers (Park et al 2008; Zhang et al 2005). These data highlight epithelial secretion-related mechanisms and the AGR genes in cancers.
Goldstein 2.59×10 \^ {-15539} \£  

[rs12686452 (C9orf72), rs12683422 (LINGO2), rs7849984 (LINGO2), ...  

Background and aims The chromosome 9p21 region encompasses several tumor suppressor genes and germline genetic polymorphisms in this region have been associated with the risk of several cancers. The goal of this study was to identify common genetic variants in the 9p21 region that were associated with multiple cancers using publicly accessible genome-wide association studies (GWAS) data. Methods We analyzed 9p21 SNPs (21,067,104-32,440,834, hg18) from eight GWAS studies with data deposited in the database of Genotype and Phenotype (dbGAP), including esophageal squamous cell carcinoma (ESCC), gastric cancer (GC), pancreatic cancer, renal cell carcinoma (RCC), lung cancer (LC), breast cancer, bladder cancer (BC), and prostate cancer (PC). The number of subjects ranged from 2,292 in PC to 7619 in LC. We first performed single SNP analyses for each study separately using logistic regression. For SNPs with P < 0.01 associated with at least one cancer, we conducted the meta-analyses through a subset-based statistical approach (ASSET) to combine P values across studies. We also calculated gene-level P values using a resampling-based adaptive rank truncated product method. Bonferroni correction was used to adjust for the number of SNPs/genes and studies examined. Results We identified five SNPs that showed significant associations for LC ([rs12668442 (C9orf172), rs12684322 (LINGO2), rs7849984 (LINGO2), rs5026895 (intergenic region of LINGO2-ACO1), and rs1052895 (intergenic region of LINGO2-ACO1)], two SNPs for ESCC (rs10631912 (CDKN2B), rs2157719 (CDKN2BAS)), two for BC (rs2764736 (intergenic region of ELAVL-TUSC1) and rs1502895 (C-T) was positively associated with meta-analysis by ASSET, rs1502895 (C-T) was positively associated with LC and BC (P = 8.16×10 \^{-15539} ).) 

Introduction: MEIS1 is a transcription factor involved in proliferation and differentiation of hematopoietic cells. MEIS1 deregulation is associated with malignancies, including leukemias. Recent studies indicate that the presence of alternative transcriptional start sites (ATSSs) in the MEIS1 gene may be related to cancer risk. SSX2 is a cancer/testis antigen normally expressed in human germ cells and several solid tumors. Conversely, SSX2 has low expression in hematopoietic malignancies. SSX2 is a nuclear protein with transcriptional repressor activity and is involved in the differentiation of hematopoietic and germ cells. In this study, we used a combination of wet-lab and in silico approaches to generate a unique list of potential ATSSs. A unique ATSS was identified at the position of the alternative translation start (positions 1218, 2230 and 3402), these possible binding sites are contained in the intronic regions 1, 3 and 4, respectively. Additionally, a possible binding site was also identified 3023 bases upstream of SSX2. This is the first report of SSX2 transcription start site deregulation. However, it is not reported in hematopoietic malignancies and leukemia-derived cell lines. We can speculate that overexpression of SSX2 represses the gene, which could compromise the repressor activity of the PolyComb group complexes. This additional study at the protein level is needed to confirm our findings.

3306W

Cryptic mutations in introns and exons leading to splice alterations in breast cancer genes. S. Casadei1, T. Walsh2, C.H. Squire1, A.M. Thornton1, J.B. Mandell1, S.M. Tray1, M.K. Lee1, M.C. King1,2, 1) Medical Genetics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA. 

Despite the widespread availability of genomic sequence, we still know very little about the frequency and spectrum of deeply intronic variants that alter splice sites and thereby influence the expression of the gene. Such mutations could lead either to extension of an exon (a bleeding exon) or to creation of a new exon from intron sequence (exonification). We are exploring this phenomenon in 594 families with multiple cases of breast cancer and/or ovarian cancer, for whom all affected relatives are wildtype for all BRCA1 and/or ovarian cancer. Our approach is to search for creation of novel cryptic exons, which can then be tested by molecular analysis. 

3307T

MEIS1 silencing influences the expression of cancer/testis antigen SSX2 in myeloid leukemia cell line K562. J. Torres-Flores1,2, A. Aguilar-Landgraf1, A. Bravo-Cuellar1, L.F. Jave-Suarez1, 1) Division of Immunology; Centro de Investigacion Biomedica de Occidente; Instituto Mexicano del Seguro Social; Guadalajara, Jalisco, 44340, Mexico; 2) Doctoral Program of Human Genetics; University of Guadalajara, Guadalajara, Jalisco, 44340, Mexico. 

Introduction: MEIS1 is a transcription factor involved in proliferation and differentiation of hematopoietic cells. MEIS1 deregulation is associated with malignancies, including leukemias. Recent studies indicate that the presence of alternative transcriptional start sites (ATSSs) in the MEIS1 gene may be related to cancer risk. SSX2 is a cancer/testis antigen normally expressed in human germ cells and several solid tumors. Conversely, SSX2 has low expression in hematopoietic malignancies. SSX2 is a nuclear protein with transcriptional repressor activity and is involved in the differentiation of hematopoietic and germ cells. In this study, we used a combination of wet-lab and in silico approaches to generate a unique list of potential ATSSs. A unique ATSS was identified at the position of the alternative translation start (positions 1218, 2230 and 3402), these possible binding sites are contained in the intronic regions 1, 3 and 4, respectively. Additionally, a possible binding site was also identified 3023 bases upstream of SSX2. This is the first report of SSX2 transcription start site deregulation. However, it is not reported in hematopoietic malignancies and leukemia-derived cell lines. We can speculate that overexpression of SSX2 represses the gene, which could compromise the repressor activity of the PolyComb group complexes. This additional study at the protein level is needed to confirm our findings.
3309W Evaluation of a cancer gene sequencing panel in a hereditary risk assessment clinic. Y. Kobayashi1, A.W. Kurian1,2,3, E. Hare4, M.A. Mills5, K.E. Kingham5, A.S. Whitemore5, V. McGuire5, G. Gong5, U. Ladabaum5, M. Cargill5, J.M. Ford5,6. 1) InVitae, San Francisco, CA; 2) Department of Medicine, Stanford University, Stanford, CA; 3) Department of Health Research and Policy, Stanford University, Stanford, CA; 4) Department of Genetics, Stanford University, Stanford, CA.

Sequencing panels of multiple cancer-associated genes are entering clinical practice, but little is known about the performance and yield of such testing among relevant patient populations. 199 adult female patients were referred to the Stanford Clinical Cancer Genetics Program for clinical assessment of hereditary breast and ovarian cancer risk; 141 had breast cancer and 57 carried known mutations in BRCA1 or BRCA2 (BRCA1/2). Germline DNA samples from these patients were sequenced on a custom multigene sequencing panel. This panel covered the entire coding region, exon-intron boundaries (±10 bp) and any other known pathogenic variants for 43 genes that have published associations with risk of breast, ovarian and other cancers, and an additional 32 cancer-associated variants elsewhere in the genome. A validation set of 200-300 samples is currently being processed. Analytic results for BRCA1/2 sequencing and pathogenicity interpretations were concordant with prior clinical testing for all patients. Twenty variants that were designated as pathogenic, either based on published literature or due to a novel truncating or splice donor/acceptor effect, were observed in genes other than BRCA1/2, including ATM, BLM, CDH1, CDKN2A, MUTYH, MLH1, NBN, PRSS1 and SLX4. Many patients carried more than one variant (including variants of uncertain clinical significance). Thirteen patients had pathogenic variants which warranted a change in cancer screening or preventive interventions based on practice guidelines; these patients were invited for confirmatory clinical testing, genetic counseling and screening recommendations. One 53-year old patient with a personal history of breast and endometrial cancers was found to carry a pathogenic MLH1 mutation; she underwent risk-reducing salpingo-oophorectomy and colonooscopy, with removal of a tubular adenoma.

Among patients referred for clinical evaluation of hereditary breast and/or ovarian cancer risk, a comprehensive sequencing assay for 43-cancer associated genes and 32 cancer-associated variants of uncertain clinical significance was evaluated. Pathogenic mutations in genes other than BRCA1/2, including ATM, BLM, CDH1, CDKN2A, MUTYH, MLH1, NBN, PRSS1 and SLX4 were identified. Pathogenic mutations were observed in 23% of patients, including variants of uncertain clinical significance. Thirteen patients were referred for confirmatory clinical testing, genetic counseling and screening recommendations. One 53-year old patient with a personal history of breast and endometrial cancers was found to carry a pathogenic MLH1 mutation; she underwent risk-reducing salpingo-oophorectomy and colonooscopy, with removal of a tubular adenoma.

3310T Novel Integrative Genomics Approach to Biomarker Discovery in Prostate Cancer. C. Hicks1,2,3, L. Meile1, T. Koganti1, J. Sethi-Amorn1, S. Vijayakumar1,1) Cancer Institute, University of Mississippi Medical Center, 2500 N. State Street, Jackson, MS 39216; 2) Department of Medicine, University of Mississippi Medical Center, 2500 N. State Street, Jackson, MS 39216; 3) Department of Radiation Oncology, University of Mississippi Medical Center, 2500 N. State Street, Jackson, MS 39216.

Recent advances in high-throughput genotyping and reduction in genotyping costs have made possible identification of common genetic variants associated with increased risk of developing prostate cancer using genome-wide association studies (GWAS). However, despite this remarkable success, several challenges remain in translating GWAS discoveries into clinically actionable biomarkers. First, genetic loci associated with prostate cancer generally explain very little of the disease risk. Second, the trait-SNP associations alone do not necessarily lead directly to the identification of causal genes and do not provide insights about the broader biological context in which the identified variants and associated genes operate. Here we report the results of a novel integrative genomics approach that combines GWAS information with gene expression data to elucidate the molecular mechanisms underpinning GWAS findings and to identify gene regulatory networks and biological pathways enriched for genetic variants. Our working hypothesis is that genes containing SNPs associated with increased risk of developing prostate cancer are functionally related and interact with one another and their downstream targets in gene networks and pathways. We tested this hypothesis using publicly available GWAS information on 300 SNPs mapped to 175 genes derived from over 350,000 cases and over 350,000 controls and gene expression data derived from 493 samples (347 prostate cancer, 65 genes associated with prostate cancer, 25 samples with metastatic cancer, and 121 control samples). We performed both supervised and unsupervised analysis followed by network and pathway analysis using Ingenuity IPA. Supervised analysis revealed 70 SNP-containing genes associated with prostate cancer and 65 genes associated with metastatic disease. Unsupervised analysis using hierarchical clustering revealed that the SNP-containing genes are co-expressed and have similar patterns of expression profiles. In addition, we identified novel genes which are functionally related with SNP-containing genes. Our novel approach using gene expression data provides insights about the broader context in which genetic variants operate.

3311F Exonic Sequencing to Identify Germline Variants in Familial Melanoma. H.N. Shabbir1,3, M.T. Landi1,2,1) Miami Dade College, Miami, FL; 2) National Cancer Institute, Department of Cancer Epidemiology and Genetics.

Germlines are the source of DNA in all cells. A mutation at the germline level is the first step to developing cancer, and the vast majority of cancer is genetic. Melanoma, the leading cause of skin cancer death, is known to be highly heritable and rare. Using a family model, high risk variants related to melanoma can be identified. The goal of the study is to integrate information from sequencing, epigenetics, and expression to identify functional and regulatory genes that are associated with melanoma. Families with two or more first-degree relatives with melanoma were considered at high risk and were investigated in this study. Initially, sequencing data of families with 3 or more relatives with the disease were examined and shared DNA variants were selected for further examination. Genetic databases and annotation tools were used to identify genes based on their known gene function and regulation, pathways, and variant conservation. Gene browsers were used to identify any histone markers, DNA methylation sites, and other epigenetic indicators. Based on our candidate genes, there is a possibility of genetic heterogeneity, in which multiple genes may be responsible for disease susceptibility. Selected candidate genes will undergo fine mapping to further investigate the region and replication in additional families and population studies of melanoma.
Patients in the State of Qatar According to World Health Organization (WHO) 2008 Criteria. This study was preliminary before further molecular investigation.

Methods: Blood samples were collected from suspected MPN cases & DNA was extracted. Allelic discrimination assays were used to evaluate point mutations causing JAK2 V617F & MPL (S505N & W515 L/K) mutations. Methods: Blood samples were collected from suspected MPN cases & DNA was extracted. Allelic discrimination assays were used to evaluate point mutations causing JAK2 V617F & MPL (S505N & W515 L/K) mutations. JAK2 Exon 12 was analyzed using High Resolution Melting Curve (HRM) assay & Sanger Sequencing. In some cases the entire MPL exon 10 & exons 12-15 was studied by RNA extraction followed by cDNA synthesis, amplification & sequencing. Results: 300 patients were tested. MPL (W515 L/K) mutations. JAK2 Exon 12 was analyzed using High Resolution Melting Curve (HRM) assay & Sanger Sequencing. In some cases the entire MPL exon 10 & exons 12-15 was studied by RNA extraction followed by cDNA synthesis, amplification & sequencing. Results: 300 patients were tested.

3313T Novel association between tanning addiction and PTCHD2 alleles using a whole-exome variant array. A. E. Bale1, A. T. Dewan1, L. M. Ferraraccio1, J. Gelernter1, D. J. Leffell2, T. T. Mayne1, J. Stapleton2, B. Carter2

Posters: Cancer Genetics

3314F The PROFILE Feasibility Study: Genetic prostate cancer risk stratification for targeted screening. R. Eiles1, E. Bancroft2, N. Taylor2, T. Dadar2, E. Page1, D. Keating1, N. Borley2, N. Desouza2, C. Goh2, M. Saunders1, G. Salter1, A. Lee1, D. Easton1, A. Antoniou2, Z. Kote-Jarai1, E. Castro1, 2

Background: Prostate cancer (PC) screening is controversial and a better assessment of individualized PC risk is needed. Several single nucleotide polymorphisms (SNPs) conferring a cumulative risk of PC have been identified. We have explored the potential role of genetic markers for targeted screening in a population with increased risk of PC due to family history (FH) of the disease. Methods: PROFILE was developed as a pilot study to determine the feasibility of targeted PC screening using prostate biopsy (PB) and its association with specific genetic profiles in men with FH. We also evaluated the role of PSA and Diffusion Weighted MRI (DW-MRI) as screening tools in this population. 100 men aged 40-69 with FH of PC were enrolled. Cumulative SNP risk scores were calculated by summing 59 risk alleles for each locus using the weighted effect (log-additive model). DW-MRI was performed in 50 patients. Participants were asked to undergo a 10 core PB regardless of baseline PSA. Results: 35% of invited men entered the study. Median age 53 years (40-69) and median PSA was 1.15. Ninety men accepted to undergo a PB as primary PC screening. Twenty-two tumours were found and 45% of them were clinically significant [Median age 64 yrs (47-69), median PSA 5.4 (0.91-9.3)] which is 24% in general population screening. The predictive performance of DW-MRI, PSA, genetic model and genetic model plus PSA measured by AUC were: 0.85, 0.73, 0.57 and 0.74, respectively. The genetic model performed best in men Analysis of a normal PSA of <3[AUC 0.63]. Analyses of a 78 SNP profile from the recent COG5 results are underway. Conclusions: Our results indicate that PB is acceptable for PC screening in men with FH of PC. The numbers of men undergoing DW-MRI were small but the AUC would warrant a larger study. The SNP risk score was more predictive in men with PSA>3 where PB would not normally be undertaken, therefore an expanded study to investigate the role of genetic profiling in directing PB in PC screening is indicated.
3315W Association between colorectal cancer risk and variants in the human exome. T.A. Harrison1, M. Lemire2, F.R. Schumacher2, P.L. Auer2, C. Ou1, S.I. Berndt3, S. Bézieau4, H. Brenner5, P. D. Campbell6, A.T. Chan7, J. Chang-Claude8, D.V. Conti8, D. Duggan7, S. Gällinger9, R. Green10, S.B. Gruber3, E. Jacobs8, S. Jiao11, L. Le Marchand12, L. Li13, P.A. Newcomb14,15, R.E. Schoen16, D. Seminara11, M.L. Slattery13, J.D. Potter17, E. White18,19, U. Peters20, G. Casey2, L. Hsu2, TJ. Hudson21,22,23,24 1) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Ontario Institute for Cancer Research, Toronto, ON, Canada; 3) Keck School of Medicine, University of Southern California, Los Angeles, CA; 4) Ziker School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI; 5) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 6) Service de Généétique Médicale, CHU Nantes, Nantes, France; 7) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany; 8) American Cancer Society, Atlanta, GA; 9) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 10) Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; 11) Translational Genomics Research Institute, Phoenix, AZ; 12) Department of Surgery, Toronto General Hospital, Toronto, ON, Canada; 13) Memorial University of Newfoundland, St. John's, NL, Canada; 14) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 15) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH; 16) Department of Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA; 17) Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD; 18) Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake UT; 19) Centre for Public Health Research, Massey University, Wellington, New Zealand; 20) School of Public Health, University of Washington, Seattle, WA; 21) Departments of Medical Biophysics and Molecular Genetics, University of Toronto, Toronto, ON, Canada.

We analyzed 5,564 cases (2,393 EA and 3,175 BE) and 10,120 controls in population-based, epidemiologic studies from the Barrett's and Esophageal Adenocarcinoma (BE) and Esophageal Adenocarcinoma (EA) Surveillance, Epidemiology, and End Results (SEER) Program and Canadian and Queensland (SEQ) Cancer registries. We performed a genome-wide association study (GWAS) using the Illumina HumanOmniExpress v1 genotyping platform. We tested 185 genes for association with CRC risk using logistic regression analysis and a Bonferroni-corrected p-value threshold of 0.05/14,570 = 3.43 × 10−6 for follow up. Using the union of all 14,570 genes, we identified 413 genes with p-values ≤ 1 × 10−4 for follow up. We integrated our results with previously published GWAS results in colorectal and gastrointestinal cancers. We found that the set of genes reported in these studies overlaps extensively, with only a few genes reaching genome-wide significance. We also found that the set of genes associated with CRC risk is highly enriched for genes associated with colorectal cancer. We conclude that our findings suggest that extensive polygenic overlap between EA and BE, our most significant results were for EA and BE combined, and, together, suggest that much of the genetic basis for EA lies in the development of BE, rather than progression from BE to EA. Further dissection of these loci is likely to lead to insights into the etiology of this rapidly fatal cancer.

3316T Pathway analysis shows different leukemia subtypes have with distinct biological mechanisms. L. Hsu1, T.B. Nggs2, R.C. Shao3, A. Chokkalingam4, K.M. Walsh3, C. Metayer5, L.F. Barcellos2, P. Buffler1 1) School of Public Health, University of California, Berkeley, Berkeley, California, CA, USA; 2) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, CA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA.

The incidence of acute lymphoblastic leukemia (ALL) has been found to be nearly 20%; higher among Hispanics than non-Hispanic Whites. However, only a few studies have been conducted in the Hispanic population. Genome-wide association studies (GWAS) have showed evidence for association in 1ZK7F, ARID5B, CEBPE, CDKN2A, and BM1-PIPK2A. However, these loci account for <10%; of leukemia genetic risk, indicating additional susceptibility loci are yet to be discovered and advanced bioinformatics tools may further guide future directions. We applied pathway-based analyses in Hispanic GWAS data of the California Childhood Leukemia Study (CCLS), testing for biological functions that are significantly enriched in childhood leukemia. Furthermore, we compared whether different biological pathways were overrepresented in major disease subtypes, including B-cell ALL, hyperdiploid B-ALL and TEL-AML1 ALL. The study population is comprised of 323 Hispanic ALL cases and 454 controls from the CCLS, using Illumina OmniExpress v1 genotyping platform. For pathway analyses, we selected genes that had at least one associated significantly SNP (P < 0.001), when adjusted for age, gender, and genetic ancestry. Of the 187 genes identified, 185 genes were incorporated for analysis using a hypergeometric test to compare the submitted list to a reference of all human genes using WebGestalt v2. The top five overrepresented KEGG pathways in ALL include axon guidance (FDR=5.1×e-06), protein digestion and absorption (FDR=0.0007), melanogenesis (FDR=0.0014), leukocyte transendothelial migration (FDR=0.0021), and focal adhesion (FDR=0.0021). Interestingly, 90% of the identified pathways are associated with cancer development, such as Wnt signaling and MAPK pathway. Between different disease subtypes, pathway analyses results indicate that hyperdiploid B-ALL and TEL-AML1 ALL involve distinct biological mechanisms compared to ALL. This is the first study to comprehensively investigate biological pathways in different leukemia disease subtypes using pathway-analyses approaches. We further used LASSO and targeted maximum likelihood estimation method to identify causal SNPs within each pathway. The results demonstrate that newly developed bioinformatics tools and causal inference methods can help identify novel loci to further understand leukemia pathogenesis.

3317F A Genome-Wide Association Study Identifies New Susceptibility Loci for Esophageal Adenocarcinoma and Barrett’s Esophagus. D.M. Levine1, W.E. Ek2, R. Fitzgerald3, S. MacGregor4, D.C. Whiteman5, T.L. Vaughan2 1) Department of Biostatistics, University of Washington School of Public Health, Seattle, WA USA; 2) Statistical Genetics, Queensland Institute of Medical Research, Queensland, Australia; 3) MRC Cancer Cell Unit, Hutchison-MRC Research Centre and University of Cambridge, Cambridge, UK; 4) Cancer Control, Queensland Institute of Medical Research, Queensland, Australia; 5) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA USA.

Esophageal adenocarcinoma (EA) is a cancer with rising incidence and poor survival. Most patients survive less than a year, and five-year survival is about 20%. For reasons that are not clear, incidence of EA is substantially higher in men and persons of European ancestry. Most EAs arise in a specialized intestinal metaplastic epithelium, which is diagnostic of EA’s precancerous lesion, Barrett’s esophagus (BE). We report the results of the first genome-wide association study of EA, and the first to examine EA together with its precancerous lesion BE. We utilized fourteen, largely population-based, epidemiologic studies from the Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON) collected over the past two decades. We analyzed 5,564 cases (2,393 EA and 3,175 BE) and 10,120 controls in two phases. For the combined case-control group we identified three novel genome-wide associations. The first is on 19p13 (P=3.6×10−10) in CRTC1, a putative cancer gene. A second is on 9q22 (P=1.0×10−9) in BARX1, which encodes a transcription factor important in esophageal specification. A third is on 3p14 (P=5.5×10−9) near the transcription factor gene, FOXP1, which regulates a transcription factor important in esophageal specification. Consistent with our previous findings showing extensive polygenic overlap between EA and BE, our most significant results were for EA and BE combined, and, together, suggest that much of the genetic basis for EA lies in the development of BE, rather than progression from BE to EA. Further dissection of these loci is likely to lead to insights into the etiology of this rapidly fatal cancer.
GENE EXPRESSION NETWORKS FOR NON-SMALL CELL LUNG CANCER MAJOR SUBTYPES. L. Liang1, 2, M.B. Freidin1, 4, E. Lim4, A.G. Nicholson2, S. Popat3, 4, M.F. Moffat1, W.O. Cookson1, 4, 1) Epidemiology, Harvard Sch Public Health, Boston, MA; 2) Biostatistics, Harvard Sch Public Health, Boston, MA; 3) National Heart and Lung Institute, Imperial College London, London, UK; 4) Department of Thoracic Surgery, Royal Brompton Hospital, London, UK; 5) Department of Histopathology, Royal Brompton Hospital, London, UK; 6) Lung Unit, Royal Marsden Hospital, London, UK.

Lung cancer (LC) remains the most common cancers worldwide for decades, and causes more than a million deaths annually. LC was in focus of numerous genomic and genomic studies including global transcriptome profiling for the purpose of identification of robust expression signatures for early diagnostics, molecular classification, disease prognosis, and response to treatment regimes. In this study we used systems biology approaches to identify transcription networks as well as individual loci that provide the way to explain transcriptome complexity of LC. Lung tumour and matched normal specimens from 69 patients with squamous cell carcinoma (SQCC, 25 patients) and adenocarcinoma (ADC, 44 patients) were analyzed. We found that gene expression profile can remarkably distinguish between tumour subtypes. Among the 36 co-expression network modules identified, three modules are highly significantly (pvalue<1e-8) associated with both subtypes, one module with ADC only and two modules with SQCC only. These networks contain very interesting patterns of genes. Controlling for false discovery rate < 0.001, we identified 2869 transcripts were differentially expressed between tumour subtypes and 69 transcripts were associated with smoking status. The hub genes of identified network modules that were consensus or specific to tumour subtypes reveal important biological pathway and molecular signature.

High-risk human papillomavirus infections are causally related to cervical cancer development. Cervical cancer represents the most dramatic cancer health disparity of women in the world, with over 200,000 deaths annually, concentrated in poor, rural, and indigenous populations. Guatemala and Venezuela are illustrative of this disparity as cervical cancer is the predominant cause of cancer cases and deaths in women. To determine the HPV viral strains in invasive tumors, we initiated a prospective collection of invasive cervical cancer cases with tissue collected in RNAlater. DNA-based viral strains in invasive tumors, we initiated a prospective collection of invasive tumors, we initiated a prospective collection of invasive tumors, we initiated a prospective collection of invasive tumors, we initiated a prospective collection of invasive tumors, we initiated a prospective collection of invasive tumors.

Materials and methods: Utilizing a large international prostate cancer genetics consortium (PRACTICAL) approximately 83,000 SNPs, enriched for prostate cancer association, genotyped in 22,924 cases and 22,889 controls were available for analysis. Prostate cancer cases were classified as having aggressive (Gleason score>7) or non-aggressive (Gleason score<7) disease. To discriminate between cases/controls and aggressive/non-aggressive disease we constructed polygenic SNP risk scores that were assessed in logistic regression models. Increasing number of SNPs were added to the score, ranked according to strength of association with prostate cancer, and evidence of biological function extracted from public resources including the ENCODE project. To select the optimal number of SNPs in the score and to avoid over fitting, a five-fold cross validation procedure was employed in a 'training dataset' (85% of data) and then evaluated in a 'test dataset' (15% of data).

Results and conclusion: Initial analysis show that prostate cancer predictive polygenic SNP risk scores including 54 established prostate cancer susceptibility SNPs result in an area under the curve (AUC) of 0.66. By adding an additional 80 SNPs into the risk score, selected by high association with prostate cancer, the predictive performance increased notably (AUC=0.70). We anticipate that this result can be further improved by incorporating functional information to rank SNPs. We show that a polygenic risk score may be a clinically useful tool to identify individuals who would benefit from targeted screening.
3321W
BRCA1 and BRCA2 (BRCA) gene analyses on an economic platform: A global consortium to demonstrate the feasibility of a shared, dedicated workflow for an non-optical next generation sequencing (NGS) with a custom BRCA AmpliSeq kit on the Ion Torrent PGM™. J. Weitzel1, J. Costa2, A. Menschenkamp3, A. Ekić4, J. Herzog5, M. Ligtenberg2, H. Fellot6, P. Park7, A. Hidalgo-Miranda8, N. Williams9, R. Ellis10, J. Carlos-Machado11, 1) Clinical Cancer Gen, City of Hope, Duarte, CA; 2) Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; 3) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Netherlands; 4) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany; 5) Department of Pathology and Molecular Medicine Queen’s University, and Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 6) Cancer Genomics Laboratory National Institute of Genomic Medicine, Distrito Federal, Mexico City, Mexico; 7) West of Scotland Genetic Services, Laboratory Medicine, Southern General Hospital, Glasgow, Scotland, UK.

While the ‘Jolie effect’ has refocused attention on the central role of BRCA gene analyses in the diagnosis and prevention of hereditary breast and ovarian cancer, there is a global disparity in access to affordable testing. The development of bench top NGS technologies holds promise for faster, more comprehensive and cost-effective methodologies than Sanger sequencing. We describe here a global consortium developing and demonstrating the feasibility of a shared, dedicated workflow for clinical grade BRCA gene analyses using Ion AmpliSeq™ multiplex PCR technology combined with Ion PGM™ System. The non-overlapping primers were designed to provide 1) 100% coverage of all coding exons and exon-intron boundaries; 2) overlapping amplicons covering exons; 3) no SNPs in the last five nucleotides of primer; and 4) maximum of three non-validated SNPs per primer. The technical protocol was developed and piloted by centers in the Netherlands and Portugal, and the current phase III testing includes 5 additional centers (California, USA; D.F., Mexico; Glasgow, Scotland; Erlangen, Germany; Ontario, Canada). More than 200 different known germline BRCA mutation positive cases were selected (~30 per center), many representative of a given region, with the aim to assess the ability to detect and call the full spectrum of mutation types, with additional cases in or within close proximity to such populations, or in less adequately characterized populations. The analyses were performed on the Ion PGM™ System at each center. Data were evaluated using Ion Reporter to facilitate use in widely disparate settings. The technical platform has been established in all 7 centers, with excellent performance characteristics: 100% coverage of the targeted regions; >100x coverage across all but 1 amplicon in BRCA2 (>40x), while using barcoding to analyze 8 cases simultaneously on a 316 chip. There was high sensitivity and specificity; technical challenges in the bioinformatic component and emerging copy number variation detection methods will be discussed. This work demonstrates the potential for mutational screening of BRCA1 and BRCA2 using the Ion AmpliSeq™ technology combined with the Ion Torrent PGM, and the feasibility of deploying a common protocol to diverse geographic settings, with a close collaboration among peers.

3322T
Inherited predisposition to cancer in Mexican women. C. M. Lankaitis1,2, A. Chaudhury1, C. Mauss2, T. Walsh3, P.A. Thompson2, A.M. Lopez1,3, A. Daneri Navarro1, M.C. King2, 1) Department of Medicine, University of Arizona, Tucson, AZ, USA; 2) University of Arizona Cancer Center, Tucson, AZ, USA; 3) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA, USA; 4) Centro Universitario De Ciencias De La Salud Universidad De Guadalajara Sierra Mojada No 560, Edificio P Primer Nivel, Colonia Independencia Guadalajara, Jalisco, Mexico.

With the emergence of comprehensive sequencing approaches for evaluation of breast and ovarian cancer genes, it is our hope to offer risk assessment and treatment for these cancers as part of an integrated program of cancer prevention for Latina women in the U.S. Southwest and Mexico. However, our experience suggests that genomic analysis of cancer risk in this population presents special challenges, because Hispanic control populations have been less thoroughly evaluated than others for naturally occurring variation in critical genes. The goal of the present pilot project is to apply BROCA capture and sequencing to identify and characterize inherited damaging mutations of all classes in all known breast and ovarian cancer genes. Participants in the study are 92 women of Mexican ancestry (49 from southern Arizona and 43 from northern Mexico) who were diagnosed with invasive breast cancer and enrolled in the ELLA Binational Breast Cancer Study. Of the participants, 15% (14/92) carry unambiguously damaging mutations (truncations, complete gene deletions, and missense mutations shown experimentally to lead to loss of protein function) in a known breast cancer gene: 5 in BRCA1, 5 in BRCA2, 2 in CHEK2, 1 in PALB2 and 1 in RAD51C. Several of these clearly damaging mutations are population specific. The challenge arises because an additional 9% of participants (8/92) carry other rare, and hence likely population-specific, mutations of unknown functional consequence in the same genes. Four participants carry variants in BRCA1 or BRCA2 at sites predicted by in silico tools to alter splice enhancers. Four other participants carry missense mutations in CHEK2 that are predicted by in silico tools to be damaging to kinase function. None of these variants appear in public databases of the 1000 Genomes Project, in dbSNP, or in the NHLBI exome sequencing project; none are characterized functionally in gene-specific databases. The problem of variants of unknown significance and population specificity is further compounded by the fact that in our experience it is disproportionately more frequent in populations with little previous evaluation for naturally occurring benign variation. Both functional and population genetics approaches will be important in resolving the meaning of this variation.

3323F
Association of Genetic variants of Cancer stem cell genes (CSC) in Gallbladder Cancer Susceptibility in North Indian population. K. Sharma1, B. Mittal2, A. Yadav1, A. Kumar2, S. Misra3, V. Kumar4. 1) Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences (SGPGIMS), Lucknow, UP, India; 2) Surgical Oncology, King George Medical University, Lucknow, India.

Introduction: Many properties of malignant cells are suggestive of those in normal stem cells. Current evidence suggests that cancer stem cells (CSC) are accountable for cancer progression and recurrence. Germline variants in CSC genes may bring about disturbed gene function and activity, thus causing interindividual variations in a patient’s chemoresistance and tumor recurrence capability. Still, the associations of genetic variations in these genes with GBC have not yet been done. Methods: The present study included 598 subjects including 398 GBC patients and 200 healthy controls from North India. This study examined association of ALCAM rs1157 G>A, CD44 rs187116 G>A, CD44 rs187115 T>C, ALDH1A1 rs13959 A>G. Genotypes were determined by self designed PCR-RFLP and ARMS-PCR. Statistical analysis was done by SPSSver16. In-silico analysis was performed using Bioinformatics tools (F-SNP, FAST-SNP, Results: Statistical analysis by logistic regression showed marginal association of ALCAM rs1157 G>A polymorphism with GBC risk (p=0.05). However CD44 rs187116 G>A, CD44 rs187116 T>C, ALDH1A1 rs13959 A>G were not associated with GBC risk. CD44 rs187116 G>A, CD44 rs187115 T>C haplotypes did not show significant association with GBC. On stratification based on gender, ALCAM rs1157 AA genotype showed increased risk of GBC in females [p=0.017] and CD44 rs187116 [GA] genotype with males [p=0.017]. Subdividing the GBC patients on the basis of gallstone status, ALCAM rs1157 G>A polymorphism imparted higher risk in patients with stones when compared to controls [p=0.04]. In case-only analysis, risk was not modified with tobacco usage and age of onset. In-silico analysis of ALCAM rs1157 G>A and CD44 rs187115 T>C revealed variable change in transcriptional and splicing regulation respectively. Conclusion: This is the first report of association of cancer stem cell variants with GBC susceptibility. The role of ALCAM rs1157 G>A with GBC risk. Grant support: DBT, DST and ICMR Government of India.
Semiconductor-based next-gen sequencing reveals cancer risk-associated genetic variations in Chiba J-MICC cohort study. J. Katayama1, Y. Hayashi1, H. Mikami2, H. Kageyama3, M. Ohira2, S. Yokoi2, H. Nagase1. 1) Life Technologies, Tokyo, Japan; 2) Chiba Cancer Center research Institute, Chiba, Japan.

Ion AmpliSeq technology is a simple, efficient and rapid process for enriching hundreds to thousands of genomic targets for next generation sequencing from 10 ng of FFPE or whole genomic DNA in a single tube. We conducted this next-gen sequencing approach to identify risk-associated genetic variants primarily in cancer risk population in a population-based cohort study following 7,900 of healthy people in a local area of Chiba since 2005. To focus the screening for cancer related gene, we used comprehensive cancer panel with Ion personal genome machine sequencer for this screening. We evaluated sufficient data amount for detection of germ-line SNVs by comprehensive cancer panel on 318chip. A preliminary study using 20 risk for stomach cancer and 8 healthy individuals by 4 barcoded run reveal over 80 percent region covered by X20 coverage. (Mean: 88.03, Range: 81.17-93.46). So, we concluded 4 samples parallel run has sufficient data amount for germ-line SNVs screening. The comparison of germ-line SNVs between risk and healthy control, we identify more frequent unknown variants and non-synonymous germ-line SNVs. These variants need the further evaluation by conventional method to avoid false positive. And also these several common variants in risk group but not in healthy group are further validating in the expanded risk and healthy population. The ability to examine 409 high-profile genes at once empowers that researchers may realize a more complete picture of the genetic variation that associated with cancer risk in a restricted population.

A support vector machine classifier for estrogen receptor positive and estrogen receptor negative early-onset breast cancer. R. Upstill-Goddard1, D. Eccles1, S. Rafiq2, W. Tapper2, J. Fliege2, A. Collins1. 1) Human Genetics and Cancer Sciences, Faculty of Medicine, University of Southampton, Duthie Building (Mailpoint 808), Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, United Kingdom; 2) Centre for Operational Research, Management Science and Information Systems, University of Southampton, Mathematics Building (54), Highfield Campus, Southampton, SO17 2JL, United Kingdom.

A major breast cancer subtype distinction depends upon the expression of receptors for the hormone estrogen on tumour cells. Tumours that express large numbers of estrogen receptors are termed estrogen receptor-positive while cancers expressing few receptors are estrogen receptor-negative. The influence of germ-line genetic variants on driving the development of either of these cancer subtypes is an important question of research. A novel approach was used to explore genome-wide single nucleotide polymorphism (SNP) data from breast cancer samples; a machine learning algorithm, called a support vector machine (SVM), was used to build a classification model capable of accurately distinguishing between estrogen receptor-positive and -negative breast cancer cases based on this SNP data. Machine learning methods have seldom been applied to genome-wide data in this way. A sample of 542 breast cancer patients, all of whom were diagnosed before the age of 40, were genotyped for ~500 000 genome-wide SNPs. A subset of 200 SNP variants was selected from the full set of SNPs based on a strong association with the receptor-negative phenotype (p<0.0005), determined using a chi-squared association test. Five SVM classification models, each using a different choice of kernel mapping function, were produced from the genotype data for the 200 SNPs in all 542 samples. In all cases classification accuracy exceeded 90%. A set of 139 genes are associated with the subset of 200 SNP variants and are thus implicated in the estrogen receptor phenotype in early-onset breast cancer. Functional classification of these genes identified enrichment for functions of the immune system. This finding is consistent with the role of inflammation in cancer and observations of a particular subtype of estrogen receptor-negative breast cancer that exhibits a good prognosis alongside extensive immune cell infiltration into the tumour site. Further exploration of the influence of SNPs on the estrogen receptor subtype distinction is on going, with particular focus on SNPs in immune system pathways. Understanding the relationship between the underlying patient genome and the process of oncogenesis is critical for the development of appropriate targeted treatment and disease prevention.
3327W

Copy number gains on chromosome 21 contribute to risk of acute lymphoblastic leukemia in non-Down syndrome children. A.J. de Smith, A. van West-Koolen, T. C. Metzler, L. F. Barcellos, P. A. Buffalo, A. Blakemore, A. P. Chokkalingam, J. L. Wielmeus. 1) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 2) Section of Investigative Medicine, Imperial College London, London, United Kingdom; 3) Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, United Kingdom; 4) Division of Neuroepidemiology, Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA 94143, USA; 5) School of Public Health, University of California Berkeley, Berkeley, CA 94704, USA.

The causes of childhood acute lymphoblastic leukemia (ALL), the most prevalent cancer in children, are largely unknown. Children with Down syndrome have an approximately 20-fold increased risk of developing ALL. The presence of trisomy 21 suggests a role for genes on chromosome 21 (Chr21). In this study, we investigated germline copy number variation on Chr21, with the hypothesis that copy number gains in non-Down syndrome individuals may be associated with increased ALL risk. Genome-wide SNP data (Illumina Human OmniExpress array) were available for 395 Hispanic children with ALL and 451 controls matched on age, gender, ethnicity and maternal race, from the California Childhood Leukemia Study (CCLS). Copy number variants (CNVs) were predicted using PennCNV, with default parameters. For quality control, samples with excess log R ratio variance and GC waviness, and CNVs with a PennCNV confidence score <20 were excluded. Cases with Down syndrome were also excluded. Association analysis was carried out in PLINK, with minimum CNV segment length at the default 20kb. We assessed the burden of large CNVs having at least one copy number gain to the genome per sample, with a cut-off of 100kb for large and 500kb for very large gains. Case-control association analysis of specific Chr21 copy number gains was carried out in samples with excess log R ratio variance and GC waviness, and CNVs were predicted using PennCNV, with default parameters. For quality control, that copy number gains in non-Down syndrome individuals may be associated with increased ALL risk.

1.44, 95% CI: 1.04-2.00, p = 0.028). The rate of very large copy number gains (>500kb) was also significantly higher in cases (0.049) than controls (0.007) (p = 6.1 x 10^-6). Common copy number gains at SLC7A91 locus were nominally associated with increased ALL risk, with 75.1% cases carrying this CNV compared to 67.6% controls (odds ratio = 1.44, 95% CI: 1.04-2.00, p = 0.028). SLC7A91 is involved in regulation of intracellular levels of folate, which is required for DNA repair and methylation, and previous studies suggest folate pathway genes may play a role in childhood ALL susceptibility. Results of this study suggest that large copy number gains on Chr21 contribute to risk of ALL in children without Down syndrome. Identifying Chr21 genes associated with childhood ALL should provide insight into the etiology of this disease, and may help to explain the increased risk of ALL in children with trisomy 21.

3329F

The cytogenetic landscape and downstream consequences in 16,172 cancer samples. L. Franke, R.S.N. Fehrmann, J. Karjalainen, L. de Vries, M. van Vugt, C. Wijmenga. 1) Department of Genetics, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands; 2) Department of Oncology, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands. With the growing public availability of gene expression data, integration and re-analysis of thousands of samples has the potential to discover new patterns that yield important novel biological insight. We integrated gene expression profiles from 77,840 samples (human, mouse and rat) to identify 2,206 transcriptional components (TCs) that explain the major regulators of gene expression. These TCs in combination with a ‘guilt-by-association’ approach allowed us to predict biological functions of individual genes, identifying and experimentally proving FEN1 as an essential homologous recombination repair gene. Many of these TCs describe metabolic, physiological and cell-type specific differences. Surprisingly, upon correcting expression data of 16,172 primary tumor samples for these TCs, we observed that the residual expression signals accurately reflected somatic copy number aberrations (SCNAs), thereby revealing the cytogenetic landscape in these tumors. A genome-wide association study at an unprecedented large scale was performed that resulted in a robust genomic signature that associated individual SCNAs to genomic instability. In addition, analysis of this cytogenetic landscape revealed clear commonalities and differences between different types of cancer. Subsequent expression quantitative trait locus (eQTL) analysis enabled us to link SCNAs to their downstream effects on biological pathways. In summary, our method provides a tool to re-analyze and re-interpret the vast amount of expression data currently available to assign functional annotation to phenotype data and to discover new biological functions of genes.

3330W

Genomic copy number signatures uncovered genetically distinct group which is different from Adenocarcinoma and Squamous cell carcinoma of the lung cancer. E. Lee, J. Lee, H. Kim, B. Shin, J. Kim, A. Kim, B. Kim. 1) Korea university Guro hospital, Seoul, South Korea; 2) Korea university College of medicine. Adenocarcinoma (AC) and squamous cell carcinoma (SCC) consisting of the most of non-small cell lung carcinoma (NSCLC) have different clinical presentation, morphology, treatment and prognosis. Recent studies suggested that fundamental genetic alterations related to carcinogenesis of each tumor type may be different. In this study, we evaluated the differences of genetic alteration of ACs and SCCs that may be applied for diagnosis of NSCLC. Genomic copy number alterations (CNAS) of 50 primary NSCLC samples (25 ACs and 25 SCCs) as well as paired normal tissue were investigated by using array comparative genomic hybridization (CGH). Common CNAs were evaluated in each subtype and compared each other to establish CNA signature. Total CNAs were more frequently identified in SCCs than ACs. The most common CNAs were gain of 3q in SCCs and 1q in ACs. Gain of 3q in SCCs was more frequently identified in SCC group which is different from Adenocarcinoma and Squamous cell carcinoma in combination of AC and SCC, which grouped genetically unique and distinct three groups by gene clustering including group III, which has different genetic signature from AC and SCC, but similar histopathology with AC or SCC. The previous study characterized genetic differences of the subtype of NSCLC including AC and SCC, which grouped genetically unique and distinct three groups by gene clustering including group III, which has different genetic signature from AC and SCC, but similar histopathology with AC or SCC. The further study of significance of this finding and the possibility of diagnostic application of this genetic marker is necessary.
3331T

The link between the losses of TP53 and ETV6 and del(5q) and/or -7/del(7q) in the progression of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) involves genetic and epigenetic changes. Targeted genomic DNA and cDNA sequencing was performed on 24 MDS/AML samples carrying del(5q) and/or -7/del(7q). Four unique combinations of CNAs are present: 1) I: cases with concurrent del(5q) and del(7q); 2) II: cases only with del(5q); 3) III: cases with concurrent del(5q) and del(7q); and 4) IV: cases with concurrent del(5q) and del(7q) and -7/del(7q). The overlap of common deleted regions (CDR) of chromosome 5 from group I and III was 5q31.1-33.1 (130,562,020-150,625,216 bp; hg18) with ~29 Mb in size and chromosome 7 from group II and III was 7q31.31-q36.1 (119,547,309-149,033,790 bp; hg18) with ~29 Mb in size. Total 249,033,790 bp: hg 18) with ~29 Mb in size. Total 249 other copy number alterations (CNAs) from the genome except chromosomes 5 and 7, which were ~78.3% of total CNAs (318), were observed. The group III (concurrent del(5q) and -7/del(7q)) is a distinctive entity carrying the most numerous other CNAs, recurrent other CNAs, cryptic CNAs (~5 Mb) and complex CNAs. Gains of PDGFD, MLL, ETS1 and ADAMTS8 and/or losses of MANF, TP53 and WRAP53 were highly associated with del(5q) since they are shared in both groups I and III. Whereas loss of ETV6 shown to be specially associated with concurrent del(5q) and -7/del(7q) (group III). Theses CNAs or genes associated with particular group or particular chromosomal abnormality may play secondary role of group specific disease progression and should be further evaluated for their clinical significance and influence on therapeutic approaches in the patient with MDS/AML carrying del(5q) and/or -7/del(7q).

3332F

Targeted genomic DNA and cDNA next generation sequencing identifies a high frequency of kinase gene fusions in Spitz tumors involving ROS1, ALK, RET, NTRK1, and BRAF. These fusions may occur as a result of oncogene activation or as a result of copy number alterations in the genomic DNA. This study used targeted genomic DNA and cDNA sequencing to identify kinase gene fusions in a series of Spitz tumors, a family of melanocytic neoplasms in which the biological behavior can range from an indolent behavior to widespread metastatic disease. The study identified a high frequency of kinase gene fusions, suggesting that these tumors may have a distinct biological behavior and that they may be more aggressive than previously thought. This study highlights the importance of identifying kinase gene fusions in Spitz tumors, as these fusions may be targets for new therapies. The study also suggests that kinase gene fusions may be more common than previously thought, and that they may be more widespread than previously thought. This study also highlights the importance of identifying kinase gene fusions in Spitz tumors, as these fusions may be targets for new therapies. The study also suggests that kinase gene fusions may be more common than previously thought, and that they may be more widespread than previously thought.
3334T
1) Laboratorio de Citogenética Genotóxicidad y Biomonitorio, Instituto de Genética Humana `Dr. Enrique Corona Rivera` BioMolecular Genomica, Univ Guadalajara, CUCS, Guadalajara, Jalisco, Mexico; 2) Unidad de Citogenética, Servicio de Hematología Oncología Pediatrica, División de Pediatría, Nuevo Hospital Civil `Dr. Juan I. Menchaca`, Guadalajara, Jalisco, Mexico; 3) Servicio de Hematología Oncología Pediatrica, División de Pediatría, Nuevo Hospital Civil `Dr. Juan I. Menchaca`, Guadalajara, Jalisco, Mexico.

Purpose: We present the cytogenetic findings in a patient with acute myeloblastic leukemia-M1 (AML-M1) at diagnosis. Clinical Presentation: We present a case of a 1 year old male, who presented tumor in the right side of the neck of approximately 8cm by 10cm size; it was multilobed, without pain and no erythema or heat. In the left side he had nodes smaller than 1 cm size. In inguinal area there was multiple nodal growth of 1 to 3cm diameter. TAC study reported neck solid lesion on the right side. Bone marrow had decreased cellularity, no megakaryocytes, 92% blasts, which were large with abundant chromatin, visible nucleoli and cytoplasm with abundant inclusions. The diagnosis was AML-M1. Method: In cytogenetic studies were performed cultures with RPMI1640 and additives, and marrow max culture media. Conventional cytogenetic study in bone marrow showed a trisomy 8, a rearrangement involving chromosomes 10 and 12, and a marker chromosome that seemed that had chromosome 11 material. For that reason Fluorescent in situ hybridization (FISH) was performed for MLL (Vysis LSI MLL Dual Color, Break Apart Rearrangement Probe), AML1/ETO (Vysis RUNX1/RUNX1T DF FISH Probe Kit), ETV6/RUNX1T1 (Vysis LSI ETV6/TELI/RUNX1AML1) ES Dual Color Translocation Probe Set, and finally performed whole chromosome painting (WCP) of 11 and 12 chromosomes (Kreatech). Results: The preliminary karyotype result was: 47,XY,+8,t(10;12)(p11.2:q11.21),-11,-mar. The FISH results were of MLL gene (nuc ish(MLLx2)(5/MLL sep 3/MLLx1)(207/300)) and an extra signal of ETO gene (nuc ish(ETOx3,AML1x2)(59/200)). And Finally the WCP results were: 47,XY,+8,t(10;12)(p11.2:q11.21),-11,-mar,del(11)(pter->p11.2::q23->p11.2::qter)\[cp30\]. Conclusions: With all previous data we concluded that there was a trisomy 8, a pericentric inversion of chromosome 11 involving MLL gene and a translocation between chromosomes 10 and 12. All this showed the preliminary karyotype result was: 47,XY,+8,t(10;12)(p11.2;21),-11,-mar. The FISH results were of MLL gene (nuc ish(MLLx2)(5/MLL sep 3/MLLx1)(207/300)) and an extra signal of ETO gene (nuc ish(ETOx3,AML1x2)(59/200)). And Finally the WCP results were: 47,XY,+8,t(10;12)(p11.2;21),-11,-mar,del(11)(pter->p11.2::q23->p11.2::qter)\[cp30\].

3335F

Genetic lesions that contribute to the progression of cancer from a primary tumor to metastases often involve a complex process of chromosome evolution, either progressive or catastrophic; however, the primary factor(s) that precede complex rearrangements remain largely unknown. As a model to investigate the mechanisms that drive metastatic potential, an inbred line of Peromyscus leucopus was isolated that carries a high frequency of malignant Harderian gland tumor formation, which immediately and predictably metastasizes to the lymphatic system, lungs and liver. Employing fluorescence in situ hybridization we have identified a recurrent abnormality, intrachromosomal amplification of Chromosome 5, among three individuals. Deep-sequencing of the microdissected Chromosome 5 has been employed to identify a candidate for the initiation of the primary lesion that contributes to both tumor susceptibility and karyotypic evolution in the metastases. Our Peromyscus strain is a highly suitable a priori model system in which to further study the mechanistic link between the derivation of distinct chromosomal aberrations in a primary tumor and those in secondary metastases.

3336W
The Impact of Microarray in Diagnosing Pediatric Acute Lymphoblastic Leukemia. C.M. Higgins, D.L. Pickering, M.L. Wiggins, W.G. Sanger, B.J. Dave. Human Genetics Laboratory, University of Nebraska Meyer Institute, University of Nebraska Medical Center, Omaha, NE.

Acute lymphoblastic leukemia (ALL) is the most common form of childhood cancer representing 23% of cancer diagnoses in children. With current therapeutic regimens, long-term survival can be expected in approximately 80% of pediatric ALL cases; however, up to 20% will remain treatment refractory. Cytogenetics and FISH have been the gold standard for defining genetic abnormalities and facilitating therapeutic stratification of pediatric ALL cases. More recently, microarray studies, which provide higher resolution and genome wide analyses, have identified novel abnormalities that are significant and recurrent with prognostic value; however, the test is yet to be included in the routine clinical diagnostic armamentarium. To better refine the diagnosis and improve prognosis/risk stratification we included microarray in our pilot studies utilizing all three techniques in six newly diagnosed pediatric B-ALL cases. We found abnormalities in all six cases with all three methodologies: cytogenetics, FISH and microarray. Ploidy changes were noted by all three techniques; known ALL-associated and prognostically significant translocations were observed by cytogenetics and FISH; the subtle translocation involving TEL/AML (ETV6/RUNX1) was detected only by FISH (2 cases); and microarray identified multiple genomic alterations that were either too subtle to be determined by cytogenetics and/or were not included in the standard ALL-FISH panel. Abnormalities determined by microarray alone included IGH deletions in 66% (4 cases), IKZF1 deletions in 50% (3 cases), deletions in EVI1 in 33% (2 cases), and a deletion of TP53 in one case. There were overlaps in detection of abnormalities between FISH and cytogenetics and microarray studies. A translocation involving the EVI1 gene region was determined by cytogenetics but could not be confirmed by FISH. Only microarray detected the microdeletion within the EVI1 region. One hyperdiploid case exhibited a deletion of TP53 which was determined by microarray and confirmed by FISH. These results demonstrate the utility of microarray in the refinement of ALL diagnosis and subsequently assist in therapeutic stratification to determine treatment regimens that are most appropriate for the disease state of these young patients.

3337T
Molecular characterization of der(1)(1;19) in a patient with myelodysplastic syndrome. J. Kim, J. Choi, J. Choi, Y. Kim, K. Lee. 1) Lab Med, Yonsei University Wonju College of Medicine, Wonju, South Korea; 2) Pathology, Yonsei University Wonju College of Medicine, Wonju, South Korea; 3) Lab Med, Yonsei University College of Medicine, Seoul, South Korea; 4) Corresponding.

Translocation (1;19)(q23;p13), creating a fusion between E2A and PBX1 genes, is a chromosomal rearrangement associated with childhood pre-B cell acute lymphoblastic leukemia (ALL), which is detected in about 5% of pediatric patients diagnosed with ALL. In contrast, rare case of der(1;19)(p13;p13.1) has been reported in mostly myeloid neoplasia or myelodysplastic syndrome (MDS). Only about 9 cases of such translocation have been reported in the literature and the specific break point has not been studied. Here, we report a case of 1;19(p13;p13.1) in a MDS patient and the characterization of the breakpoint by using conventional karyotyping, multicolor FISH and SNP array. The karyotyping revealed 47,XX,-8(8pter->p12.1::qter)(wcp13+),+der(11)(11p13.1)(19p13.1), in 20 metaphase cells, and involvement of chromosome 1 and 19 and the breakpoint was confirmed by multicolor FISH. In order to refine the breakpoint region, Affymetrix Cytoscan 750K array was carried out, and gain in the region of 1q21.2q44 (143,932,349-248,660,805) and 19p12p13.3 (260,911-24,177,726) was found, indicating a whole arm translocation. Because it is a rare chromosomal rearrangement but found recurrently in a subset of myeloid neoplasms, the possible role of such unbalanced whole arm translocation in the leukemogenesis and the possibility of gene involved in such chromosomal rearrangement need to be determined.
Characterization of Uterine Leiomyomas by Whole-genome Sequencing

M. Meire, E. Kaasinen, N. Mäkinen, R. Katainen, K. Kämpjärvi, E. Pitkänen, H. Heinonen, R. Budzow, O. Klippaara, A. Kousmanen, H. Ristolaisten, M. Gentile, J. Sjöberg, P. Vaheri, L. Aaltonen. 1) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 2) Department of Pathology, Haartman Institute, University of Helsinki, Helsinki, Finland; 3) Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland; 4) CSC-IT Center for Science, Espoo, Finland.

Uterine leiomyomas are benign but impact the health of millions of women. We performed whole genome sequencing and expression profiling of 38 leiomyomas and corresponding myometrium from 30 patients. We detected complex chromosomal rearrangements (CCRs) resembling chromothripsis in our leiomyoma samples. CCRs appear to be an unexpectedly frequent phenomenon in myometrium and had occurred in the presence of normal TP53 alleles. These rearrangements are best explained by a single event of multiple chromosomal breaks and random reassembly, and had created tissue-specific driver changes such as rearrangements between HMG2A and RADS1B loci, aberrations at the COL4A5/COL4A6 locus and interstitial deletions on chromosome 7. In some cases we found that CCRs had occurred more than once in single tumor cell lineages. We also found multiple separate tumor nodules to be clonally related. This could in part explain the common occurrence of synchronous lesions in affected uteri. Further work should examine whether the mechanisms underlying CCRs are identical to those causing chromothripsis.

Fluorescence in situ hybridization (FISH) on abnormal metaphases helps unmask the location of hidden genes/genomic segments in cases with abnormal karyotypes and abnormal FISH panels. N. Mitter, Cyto- nétics and FISH Labs, Dianon Pathology (Lab Corp), Shelton, CT.

Some complex rearrangements result in an ‘apparent’ loss of some chromosome segment or a chromosome rearrangement not detectable by classic cytogenetics analysis; but a limited FISH analysis shows the presence of two or more copies of the apparently missing chromosome segment or the splitting of a gene at a breakpoint involved in a ‘masked’ chromosome rearrangement. Microarray analysis can confirm the copy number variations for all genomic regions, but it cannot point the location of these masked/apparently missing regions. In such cases, a FISH analysis with appropriate probes on the cytogenetically abnormal looking metaphases comes to the rescue. We describe here two such cases, where in each case it was cytogenetically a real puzzle compared to the FISH panel studies, but meta-phase FISH analyses solved these puzzles. In the first case, fifteen of the twenty cells examined had only one copy of a normal chromosome 5 but also two normal signals for genes on 5p (DSS23 and DSS21). The cytogenetic analysis by itself had failed to identify the location of the second 5p. Follow-up FISH analysis on the karyotyped metaphases revealed the presence of chromosome 5 short arm material (band 5p15.2) inserted in the long arm region of chromosome 5 with a deletion of EGR1 gene at band 5q31. This implies a two-break event for chromosome 5 (most likely a pericentric inversion followed by an unbalanced translocation to the short arm of chromosome 14, resulting in presence of two copies of the probed short arm region, as well as del(5q)). This analysis, therefore, also helped in understanding the series of events that led to the unique FISH panel results. In second case, there was only one ‘normal’ copy of chromosome 3, along with presence of a long segment of extra genetic material on chromosome 11, besides other aberrations. Add(11q)mat indicated a complex banding pattern. FISH analysis for BCL6 gene revealed presence of three copies of the gene but the location of two copies was not apparent. Follow-up metaphase FISH analysis revealed presence of two copies of BCL6 gene next to each other on add(11q) material, indicating a duplication of 3q segment.

SNP Chromosomal Microarray Analysis with DNA extracted from Carney’s fixed cell pellets of Myelodysplastic Syndrome (MDS)/Acute Myeloid Leukemia (AML). N. Gentile, J. Sjöberg, O. Klippaara, A. Kousmanen, H. Ristolaisten, M. Gentile, J. Sjöberg, P. Vaheri, L. Aaltonen. 1) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 2) Department of Pathology & Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA.

SNP chromosomal microarray analysis (SNP-CMA) is an important tool in detecting very small changes in copy number, a loss of heterozygosity (LOH) in patients with inherited genetic disorders and in various cancers. The common protocol for SNP-CMA testing requires DNA extraction from fresh tissue samples. Carney’s fixed cell pellets are typically available after routine karyotyping. In our previous studies for a clinical trial, we wanted to determine the concordant discordant results from karyotype/FISH and with SNP-CMA obtained using the Affymetrix CytoScan® HD and Agilent CytoGenomics array platforms with DNA extracted from Carney’s fixed cells. For this study twelve specimens from patients diagnosed with MDS or secondary AML and with abnormal karyotype and/or FISH results, were selected. Our observations indicate that SNP-CMA is superior to karyotype/FISH in detecting LOH regions. However, routine FISH/karyotype is more sensitive than SNP-CMA for the detection of small clones, as such cases are clinically relevant for monitoring of residual/recurrence disease or its progression. Copy number changes were accurately detected with either of the array platforms when the chromosomal abnormalities are present in >30% of the cells analyzed. SNP-CMA analysis detected additional chromosomal aberrations in 4 of the 12 cases. However, these aberrations did not change the cytogenetics prognostic score. In the second part of our study, we performed SNP-CMA in 8 MDS cases (<19% blast) but with normal karyotype. FISH results were normal in all cases. There were no cases with copy-neutral loss of heterozygosity (CN-LOH) of >5 Mb, but one showed a pathogenic deletion. Our data suggest that DNA extracted from Carney’s fixed pellets is a reliable source for identification of CNVs and CN-LOH by SNP-CMA. The diagnostic yield for patients with MDS without FISH is increased by 12.5% with SNP-CMA testing. In conclusion, our study shows the utility of SNP-CMA with DNA extracted from Carney’s fixed pellets as a reflex test in patients diagnosed with MDS with a normal karyotype/FISH. This is especially useful in a clinical setting where it can prevent the need for a repeat or fresh bone marrow biopsy and the patient’s re-evaluation. Such studies will also aid in the identification of additional genomic aberrations that can have an impact on the molecular sub-classification, treatment or prognosis.

Genomic aberrations in myeloid sarcoma (MS): characterization of formalin-fixed paraffin-embedded (FFPE) samples by whole-genome SNP arrays and next-generation sequencing. M. Sukhanova, K.M. Mirza, Z. Li, F. Stoilezi, K. Oneill, M.M. Sasaki, P. Reddy, L. Joseph, G. Raca. 1) Department of Medicine, The University of Chicago Medical Center, Chicago, IL; 2) Department of Pathology, The University of Chicago Medical Center, Chicago, IL; 3) Department of Pediatrics, The University of Chicago Medical Center, Chicago, IL; 4) Medizinische Klinik und Poliklinik 1, Universitätsklinikum Carl Gustav Carus, Dresden, Germany.

Genetic aberrations in tumor cells are the strongest prognostic indicator for response to therapy and survival in patients with acute myeloid leukemia (AML). However, whole-genome sequencing of AML precursor cells localized outside of the bone marrow (so-called myeloid sarcoma), it initially does not get recognized as a hematologic malignancy, and fresh tumor samples do not get obtained for cytogenetic and molecular testing. FFPE tissue is often the only available specimen type for analysis of MS. The purpose of this study was to determine whether the analysis of FFPE samples using whole-genome SNP arrays and next-generation sequencing (NGS) allows detection of genetic abnormalities of prognostic significance in MS. We analyzed five cases of MS without BM involvement, where cytogenetic and molecular studies were either not available, or did not identify prognostically significant genetic abnormalities. Array testing was performed using CytoScan HD array (Affymetrix Inc., Santa Clara, CA). Targeted sequencing of AML associated genes was performed on the Personal Genome Machine (PGM) from Ion Torrent (Life Technologies, Carlsbad, CA), with two custom primer sets (AML and MDS panel) designed to amplify coding regions of 19 genes associated with myeloid malignancies. Somatic, disease-associated genetic aberrations were identified in all tested samples. Additional markers of significance were observed in four out of five cases. In two cases array analysis detected multiple unrelated copy number aberrations and LOH events; the presence of numerous aberrations is analogous to a complex karyotype, which is associated with a very poor prognosis. Array analysis showed an NPM1 mutation without a concomitant FLT3-ITD, a combination associated with a favorable prognosis. In conclusion, array analysis and NGS using FFPE samples revealed previously unidentified disease-associated genetic aberrations in all five MS cases, and in four cases provided information that would have been of major prognostic importance if available at the time of diagnosis. Whole-genome arrays and NGS are becoming increasingly available in the clinical setting, and should be used for detection of prognostic genomic aberrations in MS in the absence of fresh material for cytogenetic studies.
3342W
A case of pediatric acute lymphoblastic leukemia presenting with a (9;12) translocation involving JAK2 and rearrangement of MLT at 11q23 with an apparent insertion at 8q27.

Acute lymphoblastic leukemia (B-ALL) is the most common malignancy in pediatric patients and the leading cause of cancer-related deaths in children and young adults. Herein we present a 13-year-old boy who presented with abdominal pain for three months and transient fevers over the last three weeks. On presentation, he was found to have leukocytosis (WBC 76.5 × 10^3/uL), and abnormal nucleated red blood cells (abs. count 5.1 × 10^3/uL). A differential count showed 94% blasts which expressed CD34, bright CD10, CD19, partial CD20, partial CD38, partial TdT, CD79a, and HLA-DR by flow cytometry. A bone marrow biopsy was performed, and showed a hypocellular marrow extensively involved (~95%) by sheets of lymphoblasts with irregularly shaped nuclei, immature chromatin, and occasional cells with vacuolated cytoplasm. The immunophenotype of the blasts was the same as that found in the peripheral blood. A diagnosis of B-lymphoblastic leukemia was rendered. Chromosome analysis of the bone marrow showed 520 cells examined with an MLL insertion on 8q27 and a balanced translocation between 9p24 and 12p11.2. This findings were confirmed by metaphase FISH. The karyotype was described as: 46,XY,ins(6;11)(q27;p32q23);(1;12)(p24;p11.2)(5;6)XY.[15] FISH analysis using interphase nuclei also showed MLL rearrangement: MLL (11q23) split signals in 23.6% (71300) of the nuclei examined, suggestive of an MLL (11q23) gene rearrangement. Additionally, deletion of the 5^' IGH@ region, corresponding to the variable segment in 85% (265300) of the nuclei analyzed, suggestive of deletion and/or unbalanced rearrangement involving chromosome 14q32. These findings were described as: nuc ish(MLLx2)(5MLLx3)MMx1[7300] nuc ish(3IGH@ -x2.5IGHx1)[3IGH@ con 5IGH@x1][265300]. Rearrangements of 9p24 usually involve JAK2 (confirmed by FISH) induce demethylation or oligomerization of JAK2 without ligand binding, resulting in constitutive activation of JAK2, and are seen in B- Acute lymphoblastic leukemia (B-ALL) [1]. Rearrangements of the short arm of chromosome 12 are also seen in B-ALL. Correlation with other clinical and hematological data is required in cases such as this.

3344F
A Unique Rearrangement of PDGFRα and ETβV in a Three-way Translocation (t(4;12;6) in a Patient with Acute Myeloid Leukemia Progressed from Chronic Myelomycyotic Leukemia, H. M. Gürbüz, R. García, K. Wilson, S. Monaghan, P. Koduru. Det of Pathology, UT Southwestern medical center, Dallas, TX, USA.

Crypetic deletion at 4q12 leading to fusion between FIP1L1 and PDGFRA is a genetic lesion underlying myeloid neoplasms associated with hypereosinophilia. Reciprocal translocation(s) involving 4q12 leading to fusion of PDGFRA with other partners have also been infrequently reported. Here we report a unique rearrangement of PDGFRA with ETβV due to a three-way translocation t(4;12;6) in a 70-year-old man. The patient had an 8-month history of chronic myelomonocytic leukemia that was treated with high dose hydroxyurea and azacitidine. He then presented with abdominal swelling and pain and was found to have WBC of 119x109/L. Imaging revealed splenomegaly and widespread lymphadenopathy. A needle core biopsy of an abdominal mass revealed a low-grade follicular lymphoma. A bone marrow evaluation revealed acute myeloid leukemia with monocydic differentiation and low-level involvement by follicular lymphoma. The myeloid blasts expressed CD13, CD33, CD34, CD36, CD45, CD64, CD71, CD117, and CD235a. Cytogenetic analysis of bone marrow cells by G-banding demonstrated a 46,XY,t(4;12;6)(q12;p13; q21) karyotype in eighteen of the 20 analyzed cells. By FISH, there was evidence of FIP1L1/PDGFRA (4q12) gene rearrangement by both interphase and metaphase analysis which showed the PDGFRA portion of the probe translocated to 12p13. Further FISH evaluation with the ETβV break-apart probe revealed the translocation of the 5^' -end of the ETβV gene at 12p13 to chromosome band 6p21. Thus, results from both cytogenetics and FISH studies were consistent with the presence of a PDGFRA and ETβV rearrangement. Additional studies are necessary to identify the partner at 6p21 that may have fused with ETβV. The PDGFRA/ETβV fusion has previously been reported in one patient with chronic eosinophilic leukemia, who was responsive to imatinib. In AML, (4;12)(q12;p13) is recurrent abnormality with CHOPE/ETβV fusion. One prior reported case of AML with (4;12) refractory to chemotherapy was unresponsive to imatinib. Similarly, our patient did not respond to imatinib. He also failed subsequently induction chemotherapy with idarubicin and cytarabine. This is the first reported case of a three-way translocation involving PDGFRA and ETβV genes that developed into an acute myeloid leukemia without eosinophilia.

3344T

FISH analysis is superior to chromosome analysis in detecting important prognostic genetic abnormalities in PCD. However, its sensitivity is hampered due to paucity of plasma cells in whole bone marrow and often shows false-negative results when frequency of abnormal cells is below the laboratory's cut-off values. Studies have shown that the abnormality detection rate for plasma cell related abnormalities, such as secondary myelodyplastic syndrome or when diagnosis is undefined. To resolve this critical issue and optimize limited quantity received, we designed a study to develop a novel approach to detect chromosome abnormalities by sheets of lymphoblasts with irregularly shaped nuclei, immature chromatin, and occasional cells with vacuolated cytoplasm. The immunophenotype of the blasts was the same as that found in the peripheral blood. A diagnosis of B-lymphoblastic leukemia was rendered. Chromosome analysis of the bone marrow showed 520 cells examined with an MLL insertion on 8q27 and a balanced translocation between 9p24 and 12p11.2. This findings were confirmed by metaphase FISH. The karyotype was described as: 46,XY,ins(6;11)(q27;p32q23);(1;12)(p24;p11.2)(5;6)XY.[15] FISH analysis using interphase nuclei also showed MLL rearrangement: MLL (11q23) split signals in 23.6% (71300) of the nuclei examined, suggestive of an MLL (11q23) gene rearrangement. Additionally, deletion of the 5^' IGH@ region, corresponding to the variable segment in 85% (265300) of the nuclei analyzed, suggestive of deletion and/or unbalanced rearrangement involving chromosome 14q32. These findings were described as: nuc ish(MLLx2)(5MLLx3)MMx1[7300] nuc ish(3IGH@ -x2.5IGHx1)[3IGH@ con 5IGH@x1][265300]. Rearrangements of 9p24 usually involve JAK2 (confirmed by FISH) induce demethylation or oligomerization of JAK2 without ligand binding, resulting in constitutive activation of JAK2, and are seen in B- Acute lymphoblastic leukemia (B-ALL) [1]. Rearrangements of the short arm of chromosome 12 are also seen in B-ALL. Correlation with other clinical and hematological data is required in cases such as this.

3345W
Detection of the EML4-ALK fusion gene in non-small cell lung cancer (NSCLC): our FISH service to date. D. Kurczay, D. Massie, K. Kerr, D. Stevenson. 1) Cytogenticists Department, Aberdeen Royal Infirmary, National Health Service, Aberdeen, Scotland, United Kingdom; 2) Pathology Department, Aberdeen Royal Infirmary, National Health Service, Aberdeen, Scotland, United Kingdom.

The anaplastic lymphoma kinase (ALK) gene, located on 2p23, encodes a receptor tyrosine kinase involved in regulating cell proliferation and survival. Chromosomal rearrangements of the 2p23 locus can result in the production of an ALK fusion gene. One example of this is the EML4-ALK fusion gene, which is thought to be found in 3.5% of non-small cell lung cancers (NSCLC). Tyrosine kinase inhibitors (TKIs) such as Crizotinib (Xalkori, Pfizer) have been identified as potential therapies for EML4-ALK positive NSCLCs. This fusion oncogene therefore represents a significant target for novel treatments, necessitating a robust diagnostic test for evaluation of the ALK gene status in this subset of lung cancer patients. The gold standard method for identification of the EML4-ALK fusion gene is fluorescent in situ hybridisation (FISH) on formalin fixed paraffin embedded (FFPE) tissue sections. The North East of Scotland Cytogenticists laboratory has offered an ALK FISH service since March 2011, using the Vysis LSI ALK Break Apart probe. We report on our findings of a three-tiered testing strategy in conjunction with our Pathology department, which has resulted in a selected cohort of patients tested (36 diagnostic cases) with a high abnormality rate observed (76%). In addition we have also participated in the European Thoracic Oncology Platform (ETOP) Lungscape study which has involved analysing over 400 cases so far. We present our results to date, including our technical and analytical experience, as well as examples of different abnormal cases.
3346T

We describe a general workflow for the identification of tumor specific genetic variants as potential personalized therapeutic targets for the treatment of terminally ill ALL patients. Exon and transcriptome sequencing was performed to identify genetic differences between Tumor and normal tissue. One hallmark of cancer is genetic instability leading to large deletions, duplications or rearrangements, thereby generating novel fusion transcripts. These fusions very often are responsible for disease outbreak or progression. However, besides their important role in tumor biology, these fusions are found exclusively in the tumor and thus represent potential targets for treatment. In our cohort of 10 ALL patients we could identify and validate 5 gene fusion events in 4 samples. All gene fusions were copy number neutral at most likely result from balanced translocations. Three gene fusions were well-known, previously described fusion events in ALL patients. However, even in this very small cohort of 10 samples, we could identify two novel yet undiscribed gene fusions, which have a high probability of being responsible for tumorigenesis underlining the importance of individual tumor characterization.

3347F
Whole Genome Sequencing of Li-Fraumeni Families Reveals Heterogeneous Mutational Signatures. K. Bhutani1,2, D. Quarless2,3, Q. Peng2,4, K. Standish2,4, E. Scott5,6, S. Head6, S. Williams2,3, T. Kunicki2, P.B. Hedlund2, D. Nugent2, N.J. Schork1,2,7. 1) Bioinformatics Graduate Program, University of California San Diego, La Jolla, CA; 2) Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA; 3) Department of Psychiatry, University of California San Diego, La Jolla, CA; 4) The Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, CA; 5) Department of Biology, The Scripps Research Institute, La Jolla, CA; 6) Next Generation Sequencing Laboratory, The Scripps Research Institute, La Jolla, CA; 7) Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA; 8) Department of Molecular and Cellular Neuroscience, The Scripps Research Institute, La Jolla, CA; 9) Center for Inherited Blood Disorders, Children’s Hospital of Orange County, Orange, CA.

Li-Fraumeni syndrome is a rare, autosomal dominant disease caused by a germline p53 mutation. We sequenced two whole genomes (two children and both parents in each family) using Complete Genomics, Inc technology, and conducted gene expression profiling to characterize the molecular profiles of the family members. The sequence data demonstrate a dramatic increase in somatic mutation rates among carriers of p53 mutations, with Non-negative Matrix Factorization analysis revealing differing mutational signatures between carriers of zinc-finger stabilization and DNA-binding mutations. Additionally, we explored the potential of genomic analyses to influence targeted therapies for individuals using germline sequence, somatic variants, tumor sequence variants, and tumor and blood gene expression profiles.

3348W
A role for telomere length in the genetic etiology of hematological malignancies. N.B. Blackburn1, J.C. Charlesworth2,3, J.R. Marthick1, T.D. Dyer1, T.A. Thornton1, R.J. Thomson1, E.M. Tegge4,5, K.A. Marsden1,4, V. Srikanth6, J. Blangero1, R.M. Lowenthal1,4, S.J. Foote7, J.L. Dickinson1. 1) Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, Australia; 2) Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas; 3) Department of Biostatistics, University of Washington, Seattle, WA; 4) Royal Hobart Hospital, Hobart, Tasmania, Australia; 5) School of Medicine, University of Tasmania, Hobart, Tasmania, Australia; 6) Department of Medicine, Monash University, Melbourne, Victoria, Australia; 7) Australian School of Advanced Medicine, Macquarie University, New South Wales, Australia.

Haematological malignancies (HMs) are blood cancers including leukemia, lymphoma and myeloma. One of the striking features of HMs is their propensity to cluster within families indicating that genetic variation is an important risk factor for disease. Given that telomere length has been shown to be heritable in families and genetic variants are known to influence telomere length it is hypothesised that telomere length contributes, at least in part, to the familial risk of HMs. Telomere length is a known risk factor for a range of cancers with excessive telomere shortening leading to genetic instability and resulting in the malignant transformation of cells. Increasingly telomere biology is being shown to be important in the development of HMs.

Using our Tasmanian familial HMs resource, a collection of large families with multiple cases of HMs, we measured telomere length using the mono-chrome multiplex quantitative PCR method in 49 familial HM cases, 169 unaffected relatives of familial HM cases, 86 non-familial HM cases and 698 population controls. Variance components modeling of telomere length in SOLAR, taking into account familial relationships, was used to determine factors that contribute to variation in telomere length in the study samples. Through SOLAR, the heritability of telomere length was calculated to be 64.9% (P=3.2×10^{-6}, SE=0.15) indicating that around 2/3 of the variation in telomere length in these families is influenced by genetics. Adjusting for age and sex we found that overall HM cases had shorter telomeres (P=7.3×10^{-6}) than unaffected relatives and population controls and in contrast to our hypothesis this was observed across both familial (P=0.001) and non-familial cases (P=0.01). This finding supports the involvement of telomere length in the genetic etiology of both familial and non-familial HMs. Most recently we have used SNP data from Illumina 610 Quad arrays from the HM families and a genome wide quantitative trait association analysis using both the QM-QXM method and SOLAR to explore the genetic variation associated with telomere length in these families is linked to disease. Our work provides further evidence for the role of telomere length in the genetic etiology of HMs.
Identification of novel mechanisms of drug resistance in BRCA1-deficient cancer cells by exome and RNA sequencing. K.K. Dhillon1, T. Walsh2, S. Gulsuner2, T. Taniguchi1,3. 1) Divisions of Human Biology and Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute.

Acquired resistance to platinum-based therapy is an obstacle for effective treatment of ovarian and breast cancers. BRCA1 or BRCA2 loss is commonly observed in these cancers and platinum compounds, such as cisplatin and carboplatin, are initially effective for the treatment of BRCA-deficient cancers. We have shown that re-expression of BRCA1/2 due to secondary BRCA1/2 mutations in recurrent ovarian cancers is associated with platinum resistance. However, restoration of BRCA1/2 does not account for all occurrences of platinum resistance. Therefore, we hypothesize that restoration of DNA repair even in the absence of functional BRCA1/2 may lead to platinum resistance in cancer cells. To test this hypothesis, we developed an in vitro model of cisplatin resistance using the BRCA1-mutated breast cancer cell line, HCC1937. We generated cisplatin-resistant clones by culturing cells in cisplatin. Surprisingly, none of the resistant clones showed BRCA1 re-expression or secondary BRCA1 mutations. However, consistent with our hypothesis, a subset of clones restored DNA damage-induced foci formation of CtIP, RAD51 and FANCD2, which are required for DNA repair and normally require functional BRCA1. To identify mechanisms of cisplatin resistance, we performed exome and RNA-sequencing of parental HCC1937 cells and four cisplatin-resistant clones. Data from cisplatin-sensitive HCC1937 parental cells were used as a baseline to identify genetic and expression variations unique to cisplatin-resistant clones. We found that a subset of DNA repair genes is differentially expressed in cisplatin-resistant clones that have restored DNA repair foci. In the non-DNA-repair-foci-restored clone, we identified missense mutations in RAD18 and XPC, genes involved in translesion synthesis and nucleotide excision repair, which are critical for cellular resistance to cisplatin. Additionally, we observed global up-regulation of a vast majority of histone genes across all cisplatin-resistant clones, suggesting that changes in DNA packaging may contribute to cisplatin-resistance. Currently, we are conducting functional studies to assess the role and mechanism of these observed changes in cisplatin resistance. In summary, our studies showed that altered expression of DNA repair genes and histones may lead to restoration of DNA repair and cisplatin resistance in BRCA1-deficient cancer cells.

Rhabdoid tumour in a 13 year old with Ring 22: A special case of the two-hit hypothesis. H. Druker1,2,4, L. Zahavich5, D. Malkin1,3,4, S. Meyn1,2,3,5,6. 1) Cancer Genetics Program, Hospital Sick Children, Toronto, ON, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 3) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 4) Division of Oncology, Hospital for Sick Children, Toronto, ON, Canada; 5) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 6) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada.

We present a case of a severely developmentally delayed, hypotonic child with dysmorphic features and a de novo ring chromosome abnormality, 46,XX,r(22)(p11.1q13.2), who developed an abdominal rhabdoid tumour at 13 years of age. We propose a genetic mechanism and surveillance considerations for individuals with ring chromosome 22. Rhabdoid tumour syndrome is caused by germline mutations in SMARCB1, a tumour suppressor gene located at 22q11.2. Classically, young children with a germline SMARCB1 mutation have a high risk for developing atypical teratoid rhabdoid tumours of the central nervous system as well as renal and extra-renal rhabdoid tumours. Imaging is suggested for SMARCB1 mutation carriers from birth until four to six years of age for early detection and improved prognosis. Ring chromosome 22 is a rare human constitutional cytogenetic abnormality. Karyotype and FISH analyses of the patient’s blood revealed an imbalance consistent with monosomy for chromosome region 22q13.2 to 22qter but retention of the SMARCB1 gene (22q11.2) within the ring. FISH and microarray analyses of tumour tissue revealed a total loss of SMARCB1 and flanking genes in the first allele and a partial deletion of SMARCB1 in the second allele (exons 6-9). Our findings are consistent with a specific modification of the classic two hit model of tumourogenesis in which the first hit is somatic loss of the unstable ring chromosome during mitosis. In those somatic cells that have lost the ring chromosome 22, a mutation in the SMARCB1 gene on the remaining chromosome 22 (the second hit) can result in tumour development. This mechanism has been previously proposed in relation to patients with ring chromosome 22 who developed NF2 associated tumours. Increased tumour risks have also been observed in carriers of other constitutional ring chromosomes, such as r(11) and r(13) in retinoblastoma (RB1). We conclude that tumours can arise by biallelic loss of a tumour suppressor gene through the combination of loss of a ring chromosome carrying the gene and a pathogenic second mutation. Our findings suggest that patients with a ring chromosome 22 should be monitored for SMARCB1 and NF2-related tumours, and possibly CHEK2-related tumours. As the timing of the loss of the ring may be unpredictable, surveillance should be considered in these individuals beyond what is normally considered as an appropriate age cut-off.
3351W
ATM and MDC1 independently modulate resection of DNA Double-Strand Breaks for Homologous Recombination Repair in Human Cells. P. S Bradshaw1,2, M. Komosa3, M. S Meyn1,2. 1) Dept Gen & Genomic Biol 15-601F, Sick Kids Res Inst, Toronto, ON, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Canada.

Axatia-Telangiectasia (A-T) is a human multisystem genetic disorder caused by mutation in the ATM gene and characterized by progressive ataxia, immunological defects, radiation sensitivity, cancer predisposition and genomic instability. Current models suggest that the ATM protein plays a major role in the activation of the cellular DNA Damage Response (DDR) following detection of DNA Double Strand Breaks (DSBs). Establishment of a protein platform encompassing megabase regions of modified chromatin distal to the DSB thought to facilitate the ATM-mediated DDR. We find that the ATM protein is not a member of the megabase protein platform. ATM belongs to a novel class of DDR proteins whose spatial-temporal behavior following DSB induction differs from members of the megabase protein platform and from proteins directly involved in HRR. Following DNA damage induction ATM rapidly, but transiently, localizes to sites of DSBs and forms small discrete foci at damage sites that, unlike foci of H2AX and Mre11, do not diffuse across megabase chromatic regions. While MRE11 is required for ATM foci formation, ATM and MDC-1 independent MRN foci do not localize 1 hour post-irradiation, suggesting that MRN is not required for the retention of ATM at sites of DSBs. DNA damage-induced ATM foci do not co-localize with foci of the HRR-associated proteins RPA and Rad51, suggesting that ATM is associated with unresolved DSBs and may not play a direct role in HRR after DSB resection has begun. RPA ionizing Radiation Induced Foci (IRIF) in ATM- cells are of normal intensity but fewer in number. This effect is dependent on CtIP, a protein implicated in DNA resection and a target for ATM phosphorylation. In addition, depletion of CtIP in ATM+ cells results in a decrease in the number of both RPA and Rad51 IRIF and an increase in ATM IRIF. Taken together, these data support a model where ATM-mediated activation of CtIP initiates resection followed by generation of RPA-coated single-stranded DNA and the concomitant displacement of ATM from DSB sites. While loss of ATM appears to affect the proportion of DSBs that undergo resection but not the extent of resection, depletion of MDC1 increases the density of ATM IRIF without affecting their number. This result suggests that ATM and MDC1 independently modulate HRR through control of DSB resection: ATM facilitating the CtIP-dependent decision to undergo resection and MDC1 limiting the extent of resection.

3352T

Cancer-prone syndrome of premature chromatid separation with mosaic variegated aneuploidy (PCS (MVA) syndrome) is a rare autosomal recessive disorder, characterized by growth retardation, microcephaly, and childhood cancer. Mosaicism in BUB18 encoding BUBR1, a mitotic spindle assembly checkpoint regulator, has been found in individuals with the syndrome, but no second mutation was found in the second allele although a decrease of their transcript was identified. To uncover the molecular basis of the second allele, we searched for mutation in a 200-kb genomic region using a next generation sequencer. A novel single nucleotide substitution was identified in an extragenic region of BUB18 in the second allele. We then introduced it into human cultured cell lines by genome-editing technique using TAL effector nucleases (TALENs). The cell clones obtained all showed reduced BUB18 transcripts, suggesting that the nucleotide substitution is the causal mutation for the disease. Combination of next-generation sequencing and TALEN-mediated genome editing is a very powerful approach to study the molecular pathology of the disease with hitherto undetectable mutations.

3353F
Immunohistochemical Expression of Afadin-6 (AF-6) in Sporadic Neurofibromas. M. Sulaiman1, A. Dodson1, T. Hellwell1, C. Kudi1, I. Hussain1, S. Ojo1, J. Hambolu1. 1) Vet Anatomy, Ahmadu Bello University, Zaria, Nigeria; 2) Institute of Translational Medicine, University of Liverpool, UK.

Background Neurofibromatosis type 1 (NF1) - is a neurocutaneous single-gene (neurofibromin) disorder which is mapped on the long arm of chromosome 17q11.2.NF1 is inherited in an autosomal dominant manner with an incidence of about 1 in 3000. Tight junctions are specialized cell-cell point of adhesion that contributes to the regulation of differentiation and proliferation. The loss of epithelial differentiation in tumours often correlates with mutations in small GTPase Ras. The AF-6 protein is a protein that contains two potential Ras binding domains. The AF-6 functionally links the cytoskeleton through the intercellular junctions. This study was carried out to demonstrate the immunolocalization and cellular expression of AF-6 in sporadic neurofibroma by immunohistochemistry. Methods Informed patient consent was obtained before surgery and the study has an ethical approval (06/1505/137) of Liverpool Research Ethics Committee. Standard Operating Procedure of the Institute of Translational Medicine, University of Liverpool was used in the immunohistochemistry technique. Both the test and control tissues were immunostained with Rabbit Anti-AF-6 polyclonal antibody diluted at 1:100-1:200 at (pH 7.0) (Cat # 433280, Invitrogen). Slides were visualised under light microscopy. Results The AF-6 moderately localizes to the membrane of endothelial cells and at perineurial fibroblast-cell junction. Additionally, the expression of AF-6 in Schwann cells of all the Sporadic NF1 was nuclei in localization. Conclusion The study suggests that AF-6 may be involved in cell proliferation in neurofibroma cells and therefore becomes a target protein in the management of neurofibromatosis type 1.

3354W
Fragile site FS2 instability stimulates mitotic recombination in the yeast Saccharomyces cerevisiae. K. Kapellas, S. Miller, D. Rosen, E. Younkin, A. Casper. Department of Biology, Eastern Michigan University, Ypsilanti, MI.

Common fragile sites (CFS) are regions of chromosomes in mammalian cells that are prone to DNA breaks under replication stress. CFS are prevalent throughout the human genome and some lie near or within tumor suppressor genes or oncogenes. Breaks at CFS can lead to gene deletions and amplifications that can result in the genesis of cancer cells. We hypothesize that fragile site breaks also contribute to tumorgenesis by stimulating mitotic recombination events, including reciprocal crossovers (ROCs) and break-induced replication (BIR), leading to loss of heterozygosity at tumor suppressor genes. We have examined the role of fragile site instability on the stimulation of mitotic recombination using a yeast model system. The yeast Saccharomyces cerevisiae possesses the native fragile site FS2 on chromosome III. Yeast fragile site FS2 exhibits double strand DNA breaks in response to DNA replication stress, but differs in structure to mammalian CFS, being composed of simple inverted repeats. We used a galactose-inducible promoter to regulate expression of the POL1 gene of S. cerevisiae to control replication stress by galactose concentration in the medium. Three diploid yeast strains were used: an experimental strain with the native FS2 structure (AMC310), one control strain (Y382) with the FS2 region interrupted and inactivated by the insertion of a drug resistance gene, and a second control strain with the native POL1 promoter and native FS2 (AMC324). Under low-galactose conditions that cause replication stress in this strain, experimental diploid AMC310 had a RCO frequency of 15.6±10−4, a BIR frequency of 23.8±10−4, and a chromosome loss frequency of 5.7±10−4. Under no-galactose conditions, the control strain with stabilized FS2 region, Y382, exhibited a RCO rate 5-fold lower than that of AMC310, a BIR rate 12.5-fold lower than that of AMC 310, and no chromosome loss was observed. Similar results were obtained from control strain AMC324. The results indicate that FS2 is a hotspot for mitotic recombination events that lead to LOH. Future research will determine whether human CFS instability contributes to an increase in mitotic recombination and LOH in tumors.
3355T
SirT7 promotes genomic stability and adequate DNA damage response. B. Vazquez1, J. Thackray1, S. Bunting2, J.A. Tischfield3, L. Serrano1, 1) Human Genetics Institute of New Jersey Rutgers University, Piscataway, NJ; 2) Dept. of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ.

The Sir2 family of proteins, or Sirtuins, is a major player in sensing and coordinating cell stress responses by affecting cell cycle progression and mitochondrial function. Mammals have seven Sirtuins, denoted SirT1-7. They exhibit different enzymatic activities such as deacetylation, have a myriad of substrates (histone and non-histone proteins) and present a diverse pattern of cellular localization. Siruin deficiency in mice has also been associated with increased mutagenesis, defective DNA repair pathways, tumorgenesis and accelerated aging. SirT7 is the least studied mammalian Siruin. Opposing roles have been attributed to SirT7 in cancer development. Up-regulation of SirT7 expression has been observed in breast and thyroid cancer and human hepatocellular carcinomas, indicating it may be involved in the promotion and/or maintenance of oncogenic features. However, down regulation of SirT7 expression is also observed upon human and mouse cell transformation suggesting that it may have a role in cancer prevention. We investigate the role of SirT7 in the maintenance of genome integrity by characterizing SirT7 knockout mice. Our results show that SirT7-/- pups are born at sub Mendelian ratio and a high proportion died within the first month of life. In vitro mammalian cell analysis showed increased mutant frequency and polyplody in SirT7 deficient cells. Moreover, SirT7-/- fibroblasts showed increased levels of DNA double strand breaks measured by immune detection of γH2AX foci and comet assays. Furthermore, SirT7-/- thymocytes present reduced survival upon X-irradiation. B cells and T cells present impaired immunoglobulin class switching. Overall, the loss of SirT7 is associated with embryonic lethality, increased levels of spontaneous DNA damage and an impaired DNA damage response supporting the importance of SirT7 in the maintenance of genome integrity. Acknowledgments: The authors are supported by a grant from the Human Genetics Institute of New Jersey. BN Vazquez is supported by post-doctoral fellowship EX-2010-278 from the Spanish Ministry of Education, Culture, and Sports.

3356F
The Fanconi anemia pathway regulates ALT telomere maintenance in human cells. H. Root1, M. Komosa1, A. Larsen1, D.P. Bazett-Jones1, M.S. Meyn1, 1) Genetics & Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 2) Biochemistry, University of Toronto, Toronto, ON, Canada; 3) Molecular Genetics, University of Toronto, Toronto, ON, Canada.

Mutations in 15 different genes give rise to Fanconi anemia (FA), a genome instability syndrome characterized by bone marrow failure, malformations and cancer. FA proteins play roles in DNA repair, replication, meiotic recombination and telomere maintenance. The alternative lengthening of telomeres (ALT) pathway is a telomerase-independent recombination based mechanism active in cancer and immortalized human cells. We find that siRNA knockdown of FANC2, FANCA, FAN1 or FANCPC all cause rapid and dramatic increases in ALT telomeric DNA, demonstrating a central role for the FA pathway in the regulation of ALT. Depletion of FANC2 in ALT cells leads to increases in telomere length and extrachromosomal telomeric repeat (ECTR) DNA content. ECTR DNA molecules were analyzed using our novel halo-FISH technique, wherein cells are deproteinized and treated with NaOH prior to Q-FISH analysis. NaOH disrupts plectonemic interactions between ECTR DNA molecules, which, normally, DNA molecules form a closed loop. The presence of ECTR DNA at APBs primary reveals non-nucleosomal DNA, likely extrachromosomal in nature. FANC2-depletion leads to APBs that also contain chromatin intrusions, potentially representing telomeres themselves, within APBs. FISH analysis shows that in FANC2-depleted cells the frequency and stability of interactions between telomeres and ECTR DNAs is increased, suggesting that FANC2 normally suppresses interactions between ALT telomeric DNA.

FANC2 regulates a mechanism of telomeric recombination that is independent of RAD51, but requires BLM and FANCJ expression. Knockdown or inhibition of DNA-PKcs, or Ku80 does not limit telomeric DNA amplification in FANC2-depleted ALT cells, suggesting that the critical role of FANC2 in ALT is not to limit aberrant end-joining reactions. Interestingly, we also find that depletion of ATR does not result in increased telomeric DNA synthesis in ALT cells, but rather ATR expression is required to mediate the development of the ALT phenotype. This suggests that the role of FANC2 in ALT is not tied to the response to stalled/collapsed replication forks. Together, our results indicate a direct role for FANC2 and other members of the FA pathway in the regulation of recombination in the ALT pathway.

3357W
Anaplastic rhabdomyosarcoma in TP53 germline mutation carriers. S. Hetterm1, N. Archer2, G. Somers3, A. Novokmet1, A. Wagers4, L. Diller5, C. Rodriguez-Galindo1, L. Teot6, D. Malinkin7, 1) Department of Pediatric Oncology, Dana-Farber Cancer Institute and Division of Pediatric Hematology/Oncology, Boston Children's Hospital, Boston, MA 02115, USA; 2) Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada M5G 1X8; 3)Division of Oncology, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada M5G 1X8; 4) Howard Hughes Medical Institute, Department of Stem Cell and Regenerative Biology, Harvard University, Harvard Stem Cell Institute, Cambridge, MA, and Joslin Diabetes Center, Boston, MA 02115, USA; 5) Department of Pathology, Boston Children’s Hospital, Boston, MA 02115, USA.

Rhabdomyosarcoma (RMS) represents a diverse category of myogenic malignancies with marked differences in molecular alterations and histology. To determine if RMS predisposition due to germline TP53 mutations correlates with certain RMS histologies, we performed a retrospective review of RMS diagnosed in 8 consecutive children with TP53 germline mutations at Boston Children’s Hospital and the Hospital for Sick Children, Toronto. All 8 tumors exhibited non-alveolar, anaplastic histology as evidenced by the presence of enlarged hyperchromatic nuclei with or without atypical mitotic figures. Additionally, TP53 germline mutations were found in 3 out of 7 consecutive children with anaplastic RMS (anRMS) and previously unknown TP53 status. Thus, the frequency of TP53 germline mutations was 73% (11 out of 15 children) in pediatric anRMS patients. AnRMS was the first malignant diagnosis for all 11 TP53 germline mutation carriers in this cohort, and median age at diagnosis was 40 months (mean 40±15 months, range 19-67 months). The frequency of germline TP53 mutations in children with anRMS was 100% (5 out of 5 children) for those with a familial cancer history consistent with Li-Fraumeni syndrome (LFS), and 80% (4 out of 5 children) for those without an LFS cancer phenotype. Our data suggest that individuals who harbor germline TP53 mutations are predisposed to anRMS at a young age; we propose extending the Compton criteria for LFS to include children with anRMS irrespective of familial history.

3358T
Polymorphisms of MTHFR, prothrombin and factor V (Leiden) genes in children with acute lymphoblastic leukemia. E. Malý1, O. Zajac-Sypchala1, J. Kedzierska1, J. Nowak1, D. Januszkiewicz-Lewandowska1,2,3, 1) Department of Medical Diagnostics, Poznan, Poland; 2) Department of Pediatric Oncology, Hematology and Transplantation, University of Medical Sciences, Poznan, Poland; 3) Institute of Human Genetics Polish Academy of Sciences, Poznan, Poland.

The onset of thrombosis in children with acute lymphoblastic leukemia (ALL) in the course of treatment varies according to different authors, from 2.5% to 11.6% of the patients. In addition to factors related to the underlying disease and its treatment (steroids, L-ASPA, MTX, central catheter implantation), it is not without significance the presence of genetic predisposition leading to the development of thrombosis. The most common cause of congenital thrombophilia outside deficiency of protein S, C, and antithrombin, are the gene mutations of factor V Leiden, prothrombin, and MTHFR. The aim of this study was to evaluate the prevalence of polymorphisms C677T and A1298C of MTHFR, as well as prothrombin gene mutation G20210A and factor V G1691A mutation. The study was performed in 29 children (19 boys, 10 girls) treated for ALL in the Department of Pediatric Oncology, Hematology and Transplantation, University of Medical Sciences Poznan. Homozygous C677T polymorphism of MTHFR gene was found in 5 (17.2%) children. No G20210A prothrombin gene mutation was observed in the study group. The simultaneous presence of two or more polymorphisms of studied genes were found in 11 (38%) children. 3 (27%) of these patients had symptomatic thrombosis requiring treatment with unfractionated heparin. In conclusion, the frequency of homo- and heterozygous gene polymorphisms are present in children with ALL is similar to the data for Caucasians. Only the simultaneous presence of two or more polymorphisms of studied genes seems to be a predisposing factor for symptomatic thrombosis in the course of treatment of children with ALL.
Offspring of Couples Who Both Survived Cancer. J. J. Mulvihill1, J. F. Wintrrer2, L. Madanat-Harjuoja3,4, P. M. Lähteenmäki3,4, J. D. Boice Jr.3,4, 1) Departments of Pediatrics and Environmental Health, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA; 2) Danish Cancer Society Research Center, Copenhagen, Denmark; 3) Finnish Cancer Registry, Helsinki, Finland; 4) Turku University Hospital, Turku, Finland; 5) National Council on Radiation Protection and Measurements, Bethesda, MA, USA; 6) Vanderbilt University, Nashville, TN, USA.

To date and contrary to intuition and mouse studies, no environmental exposure has been proven to cause new heritable disease in human beings, not among children born to survivors of the American atomic bombs in Japan nor to survivors of cancer in childhood, adolescence, or young adulthood who received chemotherapy, radiotherapy or both. In our population-based studies of reproduction by cancer survivors in Denmark and Finland, we had the opportunity to study the birth defects, congenital anomalies and developmental disorders in offspring of that rare event when BOTH parents were cancer survivors. We identified all cases of cunnobial cancer in two population-based cancer survivor cohorts comprising 54,349 cancer survivors in the Danish or Finnish cancer registries with cancer under age 35 years between the start of the respective registries (1943 in Denmark and 1953 in Finland) through 2004, who survived to reproductive age. Offspring were linked to health registries to identify malformations, cancer, deaths (including stillbirths), selected single gene disorders and chromosomal abnormalities. A total of 58 cunnobial cancer families with a total of 110 offspring have been identified. Only 14 children were born (in 4 families) after the cancer diagnoses of BOTH parents: In 3 families each offspring had an adverse outcome (pulmonary stenosis, cleft lip and palate, and brain ependymoma, the last in a Li-Fraumeni syndrome family); in the other family (2 of 11 offspring) the patient died in infancy with an abnormality (branchiogenic cyst, terminal phalangeal aplasia of toes 1-3). In the 9 other affected offspring (born after just one parent had cancer), the father was the cancer survivor in 8 cases. No offspring has a de novo mutation. Although numbers are small, they are reassuring that therapy that is reputed to be very mutagenic in somatic cells (seen as second cancers and chromosomal breakage in the survivors) is not causing an excess of germ cell mutations.

Pleuropulmonary blastoma (PPB) type and distant metastases are significantly associated with disease-free survival: a report from the International PPB Registry. D. R. Stewart1, P. S. Rosenberg2, G. M. Williams3, J. R. Friedman4, K. A. Schultz1, A. Harris5, L. Doros6, D. A. Hill7, L. P. Dehner8, Y. H. Messinger9. 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 2) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 3) International PPB Registry and Children's Hospitals and Clinics of Minnesota, Minneapolis, MN; 4) Department of Pathology, Children's National Medical Center, Washington, DC; 5) Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO. BACKGROUND. Pleuropulmonary blastoma (PPB) is the most common primary malignancy of the pulmonary parenchyma in childhood. Three types of PPB have been recognized: type I (purely cystic), type II (cystic and solid) and type III (purely solid). Type I PPB may regress (or not progress), resulting in a cyst with the architecture of type I PPB but without the primitive cells (type Ir). Familial PPB is an autosomal dominant, pleiotropic, tumor-predisposition disorder with incomplete penetrance that arises secondary to mutations in DICER1, a gene critical in microRNA biogenesis; mutations to DICER1 result in a Li-Fraumeni syndrome family); in the fourth family, 2 of 11 offspring had an abnormality (branchiogenic cyst, terminal phalangeal aplasia of toes 1-3). In the 9 other affected offspring (born after just one parent had cancer), the father was the cancer survivor in 8 cases. No offspring has a de novo mutation. Although numbers are small, they are reassuring that therapy that is reputed to be very mutagenic in somatic cells (seen as second cancers and chromosomal breakage in the survivors) is not causing an excess of germ cell mutations.

3360W

Pleuropulmonary blastoma (PPB) type and distant metastases are significantly associated with disease-free survival: a report from the International PPB Registry. D. R. Stewart1, P. S. Rosenberg2, G. M. Williams3, J. R. Friedman4, K. A. Schultz1, A. Harris5, L. Doros6, D. A. Hill7, L. P. Dehner8, Y. H. Messinger9. 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 2) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 3) International PPB Registry and Children's Hospitals and Clinics of Minnesota, Minneapolis, MN; 4) Department of Pathology, Children's National Medical Center, Washington, DC; 5) Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO. BACKGROUND. Pleuropulmonary blastoma (PPB) is the most common primary malignancy of the pulmonary parenchyma in childhood. Three types of PPB have been recognized: type I (purely cystic), type II (cystic and solid) and type III (purely solid). Type I PPB may regress (or not progress), resulting in a cyst with the architecture of type I PPB but without the primitive cells (type Ir). Familial PPB is an autosomal dominant, pleiotropic, tumor-predisposition disorder with incomplete penetrance that arises secondary to mutations in DICER1, a gene critical in microRNA biogenesis; mutations to DICER1 result in a Li-Fraumeni syndrome family); in the fourth family, 2 of 11 offspring had an abnormality (branchiogenic cyst, terminal phalangeal aplasia of toes 1-3). In the 9 other affected offspring (born after just one parent had cancer), the father was the cancer survivor in 8 cases. No offspring has a de novo mutation. Although numbers are small, they are reassuring that therapy that is reputed to be very mutagenic in somatic cells (seen as second cancers and chromosomal breakage in the survivors) is not causing an excess of germ cell mutations.

3360F

Gene Expression in P53 and Bcl-2 in Biopsy Samples of Ulcerative Colitis and Colon Cancer in Iraqi Patients. Z. Jaafar. Biotechnology Center, Ministry of Science and Technology - Iraq, Baghdad, Baghdad, Iraq. The investigation done on the biopsy samples from patients suffered from ulcerative colitis (UC)and colon cancer. The aim of the study is to detect the genetic alteration for P53 and Bcl-2 which consider prognostic factor for chronic ulcerative colitis that show a relation between chronic ulcerative colitis and increased risk with colon cancer by using in situ hybridization technique. The results showed an increase in genetic expression for p53 and Bcl-2 in chronic UC patients as a result of accumulative mutation in P53 gene and Bcl-2 when compared with colon cancer and it can be considered as a biomarker to detect the increase sensitivity to carcinogenesis.

3361T

Improved detection of FLCN mutations in patients with Birt-Hogg-Dubé Syndrome. J. R. Toro1,2, B. Friedman1, S. Baie2. 1) National Cancer Institute, Bethesda, MD; 2) Dermatology Department, Veterans Affairs Medical Center, Washington, DC, USA; 3) GeneDx, BioReference Laboratories, Inc., Gaithersburg, MD, Maryland, USA. Background: Birt-Hogg-Dubé syndrome (BHDS) (MIM 135150) is an autosomal dominant predisposition to the development of follicular hamartomas (fibrofolliculomas), lung cysts, spontaneous pneumothorax, and kidney neoplasms. Germline mutations in FLCN are associated with the susceptibility for BHDS. To date 153 FLCN germline mutations have been reported in the online Ficolliculin sequence variation database. Objective: To characterize methods for improved FLCN mutation detection and novel mutations. Methods: Initial screening was conducted with direct bidirectional DNA sequenc- ing of the coding regions and splice sites of exons 4-14 of FLCN. If no mutation was identified by sequencing analysis, large intragenic insertion and deletion mutations were screened by RQ-PCR and targeted arrays comparative genomic hybridization with exon-level resolution. Results: The FLCN mutation detection rate by direct sequencing was 89 percent. We detected 56 unique novel FLCN germline mutations: 22 deletions, 11 insertions, 13 missense, 7 nonsense, 2 splice site and 1 deletion/insertion. To date only eight large unique intragenic mutations have been reported. We identified one whole gene FLCN deletion and eleven unique large FLCN intragenic deletions: four involving exon 1, one in exon 6 and six encompassing exons 1-6, exons 2-5, exons 2-13, exons 6-14, exons 7-8 and exons 10-14. Including this report, to date there are 214 unique FLCN mutations identified: 76 deletions, 32 insertions, 99 substitutions and 7 deletion/insertion. A comprehensive worldwide review of published FLCN mutations and clinical outcomes of patients with BHDS will be discussed. Conclusion: A systematic approach combining accurate and sensitive methods to detect FLCN mutations provides evidence that most patients with BHDS have mutations in FLCN.
3365W
Association of eNOS 4 a/b polymorphism in Mexican patients with Breast Cancer. R. Ramirez1,2, M. Gallegos1, A. Ramos1,2, L. GómezFli-
lres2, D. Carrillo1,2, O. Soto1,2, I. Gutierrez1,2, I. Delgado1, AM. Puebla1, LE. Figuera1, RP. Mariaud1, GM. Zúñiga1, 1) Laboratorio de Genética Molecu-
lar, División de Medicina Molecular, CIBO, IMSS, Guadalajara, Jalisco, Mexico; 2) Doctorado en Genética HUMana, CUCS, Universidad de Guada-
lajara; 3) Doctorado en Farmacología, CUCS, Universidad de Guadalajara;
4) División de Genética, CIBO, IMSS; 5) Laboratorio de Inmunofarmacolo-
gia, CUCEI, Universidad de Guadalajara; 6) Departamento de Clínicas Odontológicas Integrales, Centro Universitario de Ciencias de la Salud,
Universidad Guadalajara; 7) Laboratorio de mutagénesis, CIBO, IMSS.

Breast cancer (BC) is one of the most common diseases in developing
countries in the world. It is estimated that there are millions of symptomatic
women affected by BC and millions more currently asymptomatic that will
develop cancer. Several polymorphisms in the gene for endothelial nitric
oxide synthase (eNOS) have been associated with the different diseases
including cancer, one of the most studied has been the 4a/b, and however the
results have been contradiction in different parts of the world. The aim
of this study was to determine the association of eNOS 4a/b polymorphism
in BC patients. Were included samples from 429 BC patients and 281
controls from the general population, and were genotipified to eNOS 4 a/b
polymorphism in the intron 4 of gene. The allele identification was perform-
ance by polyacrylamide gel electrophoresis after staining with silver nitrate.
The observed genotype frequencies for controls and BC were 0.6% and
0.7% for a/a, 12% and 22% for a/b, 0.6% for allele b/c only control group
and 87% and 77% for b/b, respectively. The frequencies of the genotype
aa-ab of eNOS 4 a/b polymorphism showed significant differences (p <.05)
when comparing the study groups. We conclude that the genotypes a/a-a/b
of the eNOS 4 a/b polymorphism contribute significantly to breast cancer
susceptibility in the analyzed sample from the Mexican population.

3366W
Analysis of methylation pattern of candidates genes regulated by TDG in patients with germline mutations in TP53 gene. F. Fortes1, H. Kuasne2,
F. Marchi3,4, S. Rogatto3,2, M. Achatz1. 1) A.C. Camargo Cancer Center.
Department of Oncogenetics, São Paulo/SP - Brazil; São Paulo, Brazil; 2)
CIPE - NeoGene Laboratory, AC Camargo Cancer Center, Fundação Antó-
nio Prudente, São Paulo, São Paulo, Brazil; 3) Department of Urology,
School of Medicine, UNESP - São Paulo State University, Botucatu, São
Paulo, Brazil; 4) Inter-institutional Grad Program on Bioinformatics, Institute
of Mathematics and Statistics, USP - São Paulo University, São Paulo, São
Paulo, Brazil.

Li Fraumeni syndrome (LFS) is a rare autosomal dominant syndrome
caused by germline mutations in the tumor suppressor gene TP53 that
predispose to hereditary cancer. In Brazil, a LFS variant form is often due
to the occurrence of a founder effect, characterized as a mutation which
corresponds to the exchange of an arginine by a histidine at codon 337
(p.R337H mutation). It is known that p53 regulates several cellular pathways
important to cell cycle regulation, including Thymine DNA glycosylase (TDG),
a specialized glycosylase that protects the hypermethylation of CpG islands
and activates demethylation of promoters and tissue-specific enhancers.
TDG is transcriptionally regulated by p53 and is responsible for epigenetic
mechanisms which may be related to carcinogenesis in LFS.

This study was to evaluate the methylation pattern of "sentinel" genes regulated by TDG: JUN, OCT4, SOX2, HOXD8, SOX17, FOXA1, NKX2.2 and repetitive regions ALUyB8 of
gene. We analyzed methylation patterns by patients with (1) cancer and p.R337H mutation, (2) without cancer and p.R337H, (3) cancer and other mutation in TP53, (4)
relatives without cancer and without mutation in TP53, (5) controls without
cancer history and (6) 7 tumors of patients with LFS. The 5 groups were compared by pyrosequencing the gene methylation was evaluated from DNA
extracted of peripheral blood (five groups with 10 patients) and DNA
eXtracted of tumor tissue (one group with 7 LFS patients). The 5 groups were compared by patients with (1) cancer and p.R337H mutation, (2)
without cancer and p.R337H, (3) cancer and other mutation in TP53, (4)
relatives without cancer and without mutation in TP53, (5) controls without
cancer history and (6) 7 tumors of patients with LFS. It was not found any
statistical difference between the peripheral blood group, but the genes
SOX17 and HOXD8 showed altered methylation levels in the tumor group
compared to peripheral blood groups. The methylation patterns had not
changed when the tissue analyzed was blood but possibly these genes are
epigenetically regulated by other factors than TDG. Regarding to the
tumor tissue, the role of TDG in regulating HOXD8 and SOX17 methylation
cannot be assumed and it will be better assessed in other groups with and
without TP53 mutation. These results reinforce the importance of other
mechanisms which may be related to carcinogenesis in LFS.
Spliceosome components have recently been implicated as oncogenic drivers of tumor progression for a variety of human cancers. We recently reported recurrent mutations at amino acid 625 of splicing factor 3b subunit 1 (SF3B1) in ~20% of uveal melanomas. SF3B1 is a component of the U2 small nuclear ribonucleoprotein complex and the minor U12-type spliceosome. The presence of an SF3B1 mutation is a key genetic indicator of prognostic outcome in uveal melanoma. SF3B1 mutations are rarely seen in tumors that metastasized and their presence is inversely correlated with the presence of deleterious mutations in BRCA1 associated protein 1 (BAP1) that are associated with strong likelihood of metastasis. Mutations in SF3B1 have also been reported for chronic lymphoid leukemia, myeloid dysplastic syndromes, breast cancer, and other solid tumors - emphasizing the need to functionally characterize the effects of these SF3B1 mutations. To achieve this we have performed knockdown and overexpression studies of SF3B1 in cell culture followed by expression analysis with array-based approaches and RNA-Seq. Knockdown of SF3B1 in uveal melanoma cell lines from primary tumors results in the differential regulation of several oncogenes with roles in cell proliferation, angiogenesis and apoptosis. Transient overexpression of mutant forms of SF3B1 (p.R625C, p.R625H, p.R625L, and p.R625G) found in primary uveal melanomas leads to the upregulation of tumor promoting transcripts such as E2F3. No overlap of differentially regulated transcripts was observed between knockdown and SF3B1 mutant expressing lines consistent with SF3B1 mutations being gain-of-function alterations. CLK4, which regulates effectors involved in splice-site recognition was among the upregulated transcript identified in SF3B1-mutant expressing lines. Data on these cell lines are being compared with RNA-Seq data from SF3B1 mutant and wildtype uveal melanoma tumors to identify critical changes mediated by these mutations that are involved in tumorigenesis.

3369W
Functional characterization of melanoma-associated common variants in PARP1, J. Choi1, M.M. Makowski1, W.J. Kim1, T. Zhang1, M.H. Law2, M. Xu3, M. Kovacs4, H. Parikh1, L.G. Aidos5, M. Gartsdie6, H.H. Yin1, J.M. Trent1, S. Macgregor7, N.K. Hayward7, K.M. Brown1. 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Queensland Institute of Medical Research, Brisbane, QLD, Australia; 3) Translational Genomics Research Institute, Phoenix, AZ.

Recent genome wide association studies (GWAS) identified several loci for melanoma susceptibility. While these results highlight potential pathways predisposing to melanoma, functional risk variants in these regions, as well as mechanism by which they influence risk, have yet to be elucidated. To nominate potential functional variants we assessed whether SNPs in melanoma loci affect flanking gene levels by expression quantitative trait loci (eQTL) analysis. Transcript levels were measured in 62 melanoma cell lines using Affymetrix U133Plus2 expression microarrays. SNPs were then typed on Illumina OmniExpress arrays and further imputed to 1000 genomes (v3) using IMPUTE2. Regional copy number variations were also estimated and adjusted in the analysis using mach2qtl. Among 16 loci two chromosome 1 loci exhibited significant cis-eQTL. Subsequent validation using Taqman quantitative PCR (qPCR) demonstrated that Poly (ADP-ribose) polymerase 1 (PARP1) expression is significantly associated with the GWAS lead SNP (p=0.03). Namely, the risk allele is correlated with an increased PARP1 transcript levels. We then further interrogated the genotype-expression correlation by Taqman allele discrimination qPCR in 21 melanoma cell lines heterozygous for the lead SNP. The results demonstrated significantly higher proportion for the risk allele in PARP1 transcripts (p=0.001). These findings were also cross-validated using RNA sequencing on a subset of the heterozygous cell lines. Functional characterization of these SNPs demonstrates that the protective allele of rs1417765 displayed four-fold higher reporter gene expression (p=0.001). Consistent with these results, Electro Mobility Shift Assays indicated stronger nuclear protein binding for the same allele. This identification of functional SNPs in melanoma susceptibility may provide further explanation for the correlation of this SNP with PARP1 expression. Our data suggest that increased PARP1 levels are correlated with melanoma risk. Further analyses will better elucidate PARP1 function in melanoma susceptibility.

3370T
Genetic polymorphisms in DNA base excision repair gene XRCCL1 and the risk of Head and Neck Cancer. K. Chukka1, Z. Vishnuvardhan2, U. Radhakrishna1. 1) Department of Biotechnology, Acharya Nagarjuna University, Guntur, India; 2) Department of Botany & Microbiology, Acharya Nagarjuna University, Guntur, India; 3) Green cross Pathology and Molecular Biology laboratory, Paldi, Ahmedabad.

Introduction: Head and neck cancers (HNC) are among the most common types of cancer and represent a major health problem; there are approximately 540,000 new cases and 271,000 deaths annually worldwide for a mortality of approximately 50%. Single-nucleotide polymorphisms (SNPs) are the most common form of genetic variations found in the human population. The aim of this study was to evaluate the association between polymorphism XRCCL1 gene and patients with head and neck cancer. Materials and Methods: 30 head and neck cancer patients selected from the local government general hospital. An age- and sex-matched cancer-free control group (n=30) was used to compare the frequency of polymorph variants. DNA was extracted from peripheral blood. SNP were genotyped by direct sequencing. Results: Data showed no significant allelic associations for XRCCL1 347bp/445bp. 80% (n=24) of the head and neck cancer cases were Heterozygous only 20% (n=6) of the cases were Homozygous. All the controls were Heterozygous. Conclusion: XRCCL1 SNPs 347bp/445bp may not be biomarker for head and neck cancer. The complex analysis of these factors may provide the basis for personal risk assessment and an opportunity for individualized therapy.

Breast cancer is the most common cancer among women. It is estimated that 5 to 10% of the breast cancers are represented by familial breast cancers and 90 to 95% are represented by sporadic breast cancers. Mutations can lead to a change or loss of expression of a different genes and this allows the appearance of genetic and phenotypic features which contribute to tumor progression. Among these features is the ability of tumor cells to evade from the immune cells or even use immune cells in the promotion of an inflammatory microenvironment promotion which may help angiogenesis and, later, metastasis. The aim of our study was to evaluate four polymorphic variants of genes which encode important immune system molecules, two related genes encoding chemokine receptors, CCR2 (rs1799864) and CCR5 (rs333), and two related to HLA-G gene (rs1704 and rs1063320) in 105 women with familial breast cancer, 83 with sporadic breast cancer and 151 without cancer and family history of cancer (control group), such as potential markers for diagnosis and prognosis of breast cancer. Rs1799864 and rs1063320 polymorphisms were genotyped by PCR-RFLP. Rs333 and rs1704 polymorphism were genotyped by PCR. Allelic, genotypic and haplotype frequencies were estimated and compared using the Chi-square test or Fisher's exact test and subsequently were associated to diagnostic and prognostic factors. We observed a higher allelic frequency of the CCR2 wild type allele, Wt/Val (p=0.032, OR 0.46, CI 95%:0.23-0.94) and a higher haplotype frequency of the double wild type variants (Wt/Wt) of these same genes in women on the control group (p=0.030) compared to women with familial breast cancer. All polymorphisms were evaluated together with the clinical parameters and it was observed that women with breast cancer showed sporadic cancer latch (57.29±8.457 years and 44.23±12.092 years for women with sporadic and familial breast cancer respectively, p<0.001) and more invasiveness (p=0.001) as compared to women with familial breast cancer. Moreover, the rs1704 and rs1799864 polymorphism showed a positive association with tumor aggressiveness in women with sporadic breast cancer (p=0.039 and p=0.005, respectively). Our data suggest that invasive cancers increase cell infiltration and angiogenesis inflammation in the tumor microenvironment mediated by both CCR2 receptor and HLA-G molecule. Financial support: CNPq and FIPF.


Breast cancer is the most common cancer among women. It is estimated that 5 to 10% of the breast cancers are represented by familial breast cancers and 90 to 95% are represented by sporadic breast cancers. Mutations can lead to a change or loss of expression of a different genes and this allows the appearance of genetic and phenotypic features which contribute to tumor progression. Among these features is the ability of tumor cells to evade from the immune cells or even use immune cells in the promotion of an inflammatory microenvironment promotion which may help angiogenesis and, later, metastasis. The aim of our study was to evaluate four polymorphic variants of genes which encode important immune system molecules, two related genes encoding chemokine receptors, CCR2 (rs1799864) and CCR5 (rs333), and two related to HLA-G gene (rs1704 and rs1063320) in 105 women with familial breast cancer, 83 with sporadic breast cancer and 151 without cancer and family history of cancer (control group), such as potential markers for diagnosis and prognosis of breast cancer. Rs1799864 and rs1063320 polymorphisms were genotyped by PCR-RFLP. Rs333 and rs1704 polymorphism were genotyped by PCR. Allelic, genotypic and haplotype frequencies were estimated and compared using the Chi-square test or Fisher's exact test and subsequently were associated to diagnostic and prognostic factors. We observed a higher allelic frequency of the CCR2 wild type allele, Wt/Val (p=0.032, OR 0.46, CI 95%:0.23-0.94) and a higher haplotype frequency of the double wild type variants (Wt/Wt) of these same genes in women on the control group (p=0.030) compared to women with familial breast cancer. All polymorphisms were evaluated together with the clinical parameters and it was observed that women with breast cancer showed sporadic cancer latch (57.29±8.457 years and 44.23±12.092 years for women with sporadic and familial breast cancer respectively, p<0.001) and more invasiveness (p=0.001) as compared to women with familial breast cancer. Moreover, the rs1704 and rs1799864 polymorphism showed a positive association with tumor aggressiveness in women with sporadic breast cancer (p=0.039 and p=0.005, respectively). Our data suggest that invasive cancers increase cell infiltration and angiogenesis inflammation in the tumor microenvironment mediated by both CCR2 receptor and HLA-G molecule. Financial support: CNPq and FIPF.


Despite the success of genome-wide association studies (GWAS) in identifying common genetic variants associated with complex diseases, it has been difficult to demonstrate which variants are causal and how they contribute to disease risk. GWAS have previously identified a variant on 5q11.2 near MAP3K1, which is associated with breast cancer risk (rs1889312; OR = 1.13; 95% CI 1.10-1.18; p=7×10-20). MAP3K1 is a key component of the mitogen-activated protein kinase (MAPK) signal transduction pathway and is responsible for regulating important cancer genes such as c-Myc, c-Jun and c-Fos. In an attempt to determine the causal variant(s) underlying this association, we analyzed 300 genotyped and 609 imputed variants within the 305kb surrounding rs1889312 in 89,050 cancer cases in 41 case-control studies within the Breast Cancer Association Consortium (BCAC). We identified three independent association signals for estrogen receptor (ER)-positive but not ER-negative breast cancer. Using ENCODE and other published ChIP-seq data, we have identified 39-0.1kb putative regulatory elements (PREs) that are specific to each signal, and corresponding to the three association signals and each containing several candidate causal variants. Chromatin conformation studies showed long-range physical interactions between the PREs and MAP3K1, their likely target gene. The strongest signal mapped to a transcriptional silencer element in which the risk alleles of candidate causal variants increased MAP3K1 promoter activity in luciferase reporter assays. Electrophotoptive mobility shift assays (EMSA)s have detected allele-specific protein-DNA interactions for several candidate causal variants. In conclusion, we have identified candidate single nucleotide variant (SNV) associations at 5q11 that are likely to be causally related to breast cancer risk and act by controlling MAP3K1 expression.
Drug-sensitivity assay for sequence variants verification. K. Tao, S. Tavtigian. Huntsman Cancer Institute, Salt Lake City, UT. [Background] Individual sequence variant may confer different level of drug sensitivity and affect the outcome of treatment. Emerging poly (ADP-ribose) polymerase (PARP) inhibitor-based synthetic lethal approach is beneficial for the patients with defective function in homologous recombination repair system. High-throughput sequencing technologies provide us numerous sequence variants and majority of them are missense substitutions that are difficult to interpret. However, the assay to determine which sequence variant create clinical significance or increase drug sensitivity has not been well established. Here, we developed a drug-sensitivity assay system to verify sequence variants of genes involved in HR pathway. [Method] Total twelve sequence variants were analyzed in this method. Stable cell lines that have either a wild-type or mutant construct were generated using PiggyBac (PB) transposase (PB 5’ terminal repeat, EF1alpha promoter, COSHbox, IRES_RFP, PB 3’ terminal repeat). For accurate and high-throughput screening, the cells expressing the same level of fluorescence intensity were sorted and plated in 96 well plates using FACSAnia cell sorter. Cross-linking agent (Mitomycin, Cisplatin), poly ADP-ribose polymerase (PRRP) inhibitor (AZD-2281, MK4827), and Topoisomerase I (Camptothecin) were used individually or as combination. Following the drug exposure for 5-7 days, cell viability was assessed based on the amount of APT. The APT-generated luminescent signal of each well was measured in EnVision 2104 plate reader. [Result] Three sequence variants (5’UTR variant, missense substitution, and frameshift deletion) showed an increase sensitivity to the drugs compared to wild-type. These three mutants were sensitive to both cross-linking agent and PARP inhibitor and the combination of these two agents augmented drug resistance. [Conclusion] It is critical to screen and prioritize treatment for individual cancer patient and for that, the development of system for accurate measurement of drug sensitivity in each sequence variant is urgent. Our strategy allows to see the effect of one-week treatment, which is necessary to determine the effect of PARP inhibitor and three mutants were confirmed as pathogenic variants. However, pathogenicity of the remaining mutants are not yet clear. Further improvement of sequence verification system is necessary to detect smaller functional difference and it is ideal to conduct the assay using isogenic human mutant cells.

SNP Variation in MicroRNAs Targeting the Tumor Suppressor Gene PTEN. V.A. Ware, A. Jones, C. LaViolette, C. Corcoran, A. Maletz, J.A. Wilder. Northern Arizona University, Flagstaff, AZ. MicroRNAs play important roles in a broad range of biological processes and previous studies have demonstrated that SNP variation in microRNA sequences can affect pathways associated with cancer progression through diverse mechanisms. Here we describe sequence variation in microRNAs that target PTEN (Phosphatase and Tensin Homolog), an important tumor suppressor gene in a number of cancers including prostate, breast and lung cancer. We screened 41 Native American samples to find SNPs in microRNA encoding genes unique to Native American populations. A SNP (rs13136737) was found in the pri-miRNA of mir302D and mir367, which we are evaluating via quantitative PCR for potential allele-specific patterns of expression. Further evaluation of microRNA-encoding genes may uncover variation affecting cancer progression.

Missense Mutation of the Last Nucleotide of Exon 1, CDH1 c.48 G>C (Q16H), Contributes to Cancer Predisposition through Disruption of Normal Splicing and Generation of Missense Mutation. L. Zhang1, A. Xiao2, J. Ruggen1, R. Bacares3, S. Melo3, J. Figueiredo1, J. Simões-Correia4, R. Seruca1, M. Shah4. 1) Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Northwestern University, Evanston, IL; 3) IPATIMUP - Instituto de Moléculas Patológicas and Instituto de Tecnologia da Universidade do Porto, 4200-465 Porto, Portugal; 4) Department of Medicine, Weill Cornell Medical College, New York, NY. Purpose: The objective of this analysis was to assess the molecular mechanisms through which the CDH1 c.48 G>C (Q16H) missense mutation contributes to cancer predisposition. Mutation screening of CDH1 that contributes to the Hereditary Diffuse Gastric Cancer (HDGC) is important for patients with early-onset gastric cancer and/or with strong family history. Classification and prediction of CDH1 critical for cancer predisposition management. Particular challenge is the classification of rare non-truncating CDH1 sequence variants because it is not known whether these subtle changes can affect E-cadherin protein function sufficiently to predispose cells to cancer development. Patients and Methods: Our case is a 22 year old woman who was diagnosed with gastric cancer at the age of 18. Her grandmother, one great aunt and one great uncle were diagnosed with gastric cancer in their 30s and died of this disease in their 30s or 40s. The other great aunt was diagnosed with abdominal cancer at 16 and died subsequently. The CDH1 c.48G>C variant was identified in the patient and her living great uncle. RT-PCR products from the patient and her great uncle were cloned into the TOPO vector and subsequent sequencing was performed on individual clones from these two patients. In vitro functional studies were performed by transient transfection of CHO cells with the vector encoding the WT and the mutant protein, as well as the empty vector (Mock), as a control. Upon transfection, cells were analysed for E-cadherin expression and localization and more importantly, for the two main functions of E-cadherin: cell-cell adhesion and invasion suppression. Results: Using RT-PCR and subsequent cloning strategy, we were able to detect a low level of mutant transcripts (5/27 clones had the ‘C’ allele, Q16H) in our patient. However, we were unable to detect any mutant transcripts in her great uncle (0/17 clone had Q16H). In vitro functional studies of the Q16H mutation demonstrated it did not affect the expression, localization and cell-cell adhesion function of the E-cadherin. Instead, we observed loss of anti-invasion function in the E-cadherin protein with this substitution. Conclusion: These results support the conclusion that results that CDH1 c.48G>C (Q16H) variant is a deleterious mutation and contributes to HDGC through disruption of normal splicing and inducing increased cell invasion.
Detection of P16, Cyclin D1 and Bcl-2 Expression in Pancreatic Neoplasms. S. Lai1,2, X. Zhou1. 1) Pathology, Michael E. DeBakey VA Medical Center, Houston, TX; 2) Baylor College of Medicine, Houston, TX.

P16 and cyclin D1 genes play a critical role in the regulation of the G1-S transition of the cell cycle, and their expression are frequently altered in several neoplastic entities. Bcl-2 gene in apoptosis pathway has important function in tumorigenesis of cancers. Analysis of the protein products of these genes by immunohistochemical methods provides information on their functional status and allows for their phenotypic evaluation of tumor cells. Studies of the correlation between these gene expression levels and clinicopathologic characteristics of pancreatic neoplasms are not ample. We constructed a tissue microarray block from 24 pancreatic neoplasms including 17 adenocarcinoma, 6 neuroendocrine neoplasm and 1 intraductal papillary mucinous neoplasm, and stained for P16, cyclin D1 and Bcl-2 by immunohistochemistry. Protein expressions were correlated with tumor histologic type, grade, perineural invasion, lymphovascular invasion and lymph node status. Loss of P16 expression was correlated with malignancy (p<0.001), and lymphovascular invasion (p=0.032). Cyclin D1 overexpression was also associated with malignancy (p=0.035). Our study indicates the absence of p16 in most of the malignant neoplasms are associated with an early tumorigenic event. High P16 expression level in pancreatic tumors predicts a better prognosis. P16 and cyclin D1 immunoreactivity may be used as a prognostic marker for pancreatic neoplasms.

Exome-based method to determine cancer tissue of origin. K. Robasky1, E. Aronesty1,2, W.D. Jones1,2. 1) Expression Analysis a Quintiles Company, Durham, NC; 2) Bioinformatics Program, Johns Hopkins University, Washington D.C.; 3) Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC.

Identifying the source of newly presenting tumor fails to yield definitive results in a significant number of cases, with cancer of unknown primary organ (CUP) accounting for 3-5% of all cancer diagnoses. Tumors are commonly classified via visual inspection alongside of immunohistochemistry and expression profiling, neither of which can consistently identify a molecular signature by which to guide treatment. Misclassifying a cancer’s primary site can have wide-ranging effects, including: 1) masking the true primary cancer, 2) misinforming the patient treatment plan 3) obfuscating clinical trial results by enrolling misclassified subjects. Here we present a method for using cost-effective, high-throughput sequencing exome data to find the molecular signatures of various tissue-specific cancers. We do so by first scoring the genes from individual exomes based upon the level of mutational burden. We then perform an unsupervised analysis to reduce the dimensionality of the scored genes. Using exomes from The Cancer Genome Atlas, we apply this method to classify cancers into tissue-specific clusters. This model can thus be used with new tumor exomes to aid in identifying the tissue-of-origin of the primary cancer.

A novel statistic method for drug response prediction with big RNA-seq data. M. Xiong1, L. Ma2, M. Chen1, S. Guo2,3. 1) Dept Biostatistics, Univ Texas Hlth Sci, Houston, TX; 2) Fudan University, Shanghai, China.

Digital transcriptome analysis by next-generation sequencing discovers substantial mRNA variants. Variation in gene expression underlies many biological processes and holds a key to unraveling mechanism of common diseases and drug response. However, the current methods for drug response prediction using overall gene-expression are originally designed for microarray expression data and overlook a large number of variations in gene expressions. The challenge for application of gene expression to pharmacogenomics is how to accurately measure variation in gene expression and how to develop statistical methods to fully explore information on expression variation for predicting response of a patient to a particular treatment regimen with extremely big RNA-seq data. To fully utilize expression information at genomic positional level we use sufficient dimension reduction (SDR) techniques which project the original high dimensional data to very low dimensional space while preserving all information on response phenotypes, functional principal component analysis which further reduce the dimension of the big RNA-seq data, and penalization methods for optimal feature selection as a powerful tool to develop novel statistical methods for classifying drug response of patients with RNA-seq data. The proposed methods are applied to ovarian cancer drug response RNA-seq and microarray expression data from TCGA database where expressions of 163 sensitive and 70 resistant to drug ovarian cancer samples were measured by RNA-seq and microarray. Tenfold cross validation was used to evaluate the performance of classifier. We can reach almost 100%, 100% and 100% average classification accuracy, sensitivity and specificity, respectively, in the test datasets with RNA-seq data, and 63.54%, 88.85% and 10.43% average classification accuracy, sensitivity and specificity, respectively, in the test datasets with microarray-measured expression dataset. Our results strongly demonstrate that the prediction based on RNA-seq substantially outperforms the prediction based on the microarray expressions and open a new avenue for cancer pharmacogenomics studies.

Leukemia relapse in donor cells ten years after allogeneic hematopoietic stem cell transplantation for acute myeloid leukemia. G. Calabrese1,2, L. Milioti1, R. DiGianfilippo1, M. Alfonsi1,2, P. Guanciali-Franchi1, D. Fantasia1, P. Bavaro3, P. DiBartolomeo3, G. Palka1,2,1. 1) Dept S Med, Oral & Biotec, Univ Chieti, Chieti Scalo, Italy; 2) Medical Genetics Dept, Pescara Hospital, Pescara, Italy; 3) Bone Marrow Transplantation Center, Pescara Hospital, Pescara.

Leukemic relapses after allogeneic hematopoietic stem cell transplantation (HSCT) in most instances arise from the original recipient leukemia clone. Exceptionally, leukemic relapse can originate from the donor hematopoietic cells. A recent review listed 65 cases of donor cell leukemia, with an estimated risk to be about 12 per 10,000 HSCT. Exposure to radiant therapy and to chemotherapy are regarded as involved in donor cell leukemogenesis, though causes are still unclear. In 2000, a female patient, 42-ys-old, was diagnosed with acute myeloid leukemia (AML) M2 subtype having a diploid karyotype. After achieving remission by induction chemotherapy and total body radiotherapy the patient underwent a HSCT with cells from her brother. Six months later the patient showed a complete chimerism with a normal male karyotype which persisted for nine years. In 2010 the patient was treated with metabolic radiotherapy for a thyroid carcinoma. Six months later showed pancytopenia and FISH analysis with XY chromosome probes revealed a partial chimerism (0.08% recipient cells). Six months after loss of complete chimerism the patient showed clinical-hematologic features of leukemia relapse. Cytogenetic analysis performed on bone marrow aspirate revealed an abnormal male karyotype with chromosomes 8, and 13 clonal abnormalities, and including a marker chromosome in all cells analyzed. This result was confirmed by FISH. The patient deceased two weeks later. In the present case AML occurred in donor cells ten years after HSCT with a phenotype overlapping patient original disease. Although rarely reported, the chromosomal anomalies in donor cells were specific of therapy-related AML. Radiotherapy for thyroid carcinoma and possible immune system disturbance might be associated with leukemogenesis. Chimerism study with the finding of clonal chromosomal anomalies confirm cytogenetic analysis role as of relevance for both HSCT follow up surveillance and investigation of biological mechanism of leukemogenesis.
3383F
Photodynamic therapy effect in EGFR pathway gene expression in glioblastoma as a biological model. L.B. de Paula1,2, F.L. Primo1,2, N.T.A. Pereira1,3, H.R. Kazmi2,4, A. Rossi2,4, A. Chandrasekaran1,2,4. 1) Center of Nanotechnology and Tissue Engineering - Photobiology and Photomedicine research Group, School of Philosophy, Sciences and Letters of Ribeirão Preto, University of São Paulo, São Paulo, Ribeirão Preto, Brazil; 2) Department of Genetics, Medical School of Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo, Ribeirão Preto, Brazil; 3) Department of Chemistry, School of Philosophy, Sciences and Letters of Ribeirão Preto, University of São Paulo, São Paulo, Ribeirão Preto, Brazil; 4) Laboratory of Genetics and Molecular Biology of Fungi, Medical School of Ribeirão Preto and University of São Paulo, São Paulo, Ribeirão Preto, Brazil.

Glioblastoma multiforme (GBM), the most malignant human brain tumor, may develop through high pressure necrosis or apoptosis, thereby causing astrocystoma. The deadly nature of GBM originates from explosive growth and invasive behavior, which are fueled by dysregulation of multiple signaling pathways. Dysregulated epidermal growth factor receptor (EGFR) pathway is frequently in high-grade gliomas via gene amplification or dominant-active mutation, making it a leading cause of gliomagenesis. New methods to achieve widespread distribution of therapeutic agents have shown a significant improvement in brain tumor therapy. Photodynamic therapy (PDT) has been used mainly as an anticancer therapy, which relies on the absorption and retention of a photosensitizer molecule in the tumor cells associated with further irradiation with an appropriate visible light. We investigated whether similar changes in gene expression EGFR pathway in GBM cell lines pre and post treatment with PDT could be detected and correlating the incorporation of this molecule as a treatment for different glioma stages.

Gene expression EGFR pathway resulted in hipoexpression EGFR, SOS1, SOS2, STAT1, PIK3R1, PIK3CA, PIK3CB, MAP2K4, MAP3K1, MAPK, MAPK10 and MAPK6 MAPK9 in the post-PDT. The complex 3 (MAP3K/MEK/ERK) plays a key role in many processes that affect tumor growth and progression, including proliferation, dedifferentiation, apoptosis inhibition, invasiveness and loss of dependence on adhesion. Phosphorylation of tyrosine kinase residues in EGFR functions as binding domain of Grb2/Sos complex, which activates the signaling cascade Ras/Raf/MEK/MAPK pathway, which influences the proliferation, migration and differentiation. Members of the MAPK pathway have been associated with various functions within the process of tumorigenesis, since this pathway is responsible for the regulation of proliferation, differentiation, survival, angiogenesis and metastasis in many tumor types. These findings confirmed that the engineering of nanocarriers associated with PDT procedures led to hipoexpression of the genes that are directly involved in the tumor process the EGFR pathway in vitro by advanced protocols that can be useful for future in vivo trials available to clinical oncology. With the development of protocols associating photodynamic therapy and cancer, we aim to develop advanced treatments with nanotechnology tools for highly resistant diseases such as human gliomas.

3385T
Prognostic signature in papillary thyroid carcinoma patients that undergone total thyroid ablation. M.C. Barros Filho1, F.A. March2, C.A. Renato1, S.R. Rogato1,2, L.P. Kowalski1, 1) AC Camargo Cancer Center, Sao Paulo, Brazil; 2) Sao Paulo University, Sao Paulo, Brazil; 3) Faculty of Medicine/UNESP, Botucatu, SP, Brazil.

Purpose. Papillary thyroid carcinoma (PTC) is the most common endocrine malignancy. Despite the low mortality rate, recidive is frequent and associated with patient morbidity. Currently, clinical and pathological procedures are insufficient to determine the risk of relapse. Although BRAF mutation is often associated with tumor aggressiveness, no molecular marker is used in clinical routine. The aim of this study is to identify molecular markers able to predict recurrent in PTC patients. Patients and Methods. Eighty-three patients submitted to total thyroid ablation were retrospectively enrolled. Primary fresh-frozen tumors (RNA integrity score> 7) from 58 cases were evaluated by expression oligoarray using Sure Print G3 8x60K slides (Agilent Technologies). Cases without suspicion of active disease (normal imaging screening and serum Tg) with at least five years of follow up (group 1= 46) were statistically compared with cases with confirmed recurrent disease (group 2= 12) by SAM (FDR< 0.05) and t test (P< 0.005). Six transcripts were selected for further RT-qPCR confirmation in a technical validation group (samples previously evaluated by microarray) and in an independent validation group (group 1= 15 and group 2= 10). Results. By expression oligoarray, two genes were underexpressed (FBXL5 and F2RL2) and 13 overexpressed (FOXP2, SLC24A, GADD45B, C6ORF22, LYSM3D, UTP23, C1ORF79, NCRNA00203, FLJ45950, S1PR1, DNTT, MB2D and PROM2) in the recurrence group. Four out of six transcripts evaluated by RT-qPCR were highly correlated with microarray data (Pearson correlation >0.98). The other 13 genes had a lower correlation with recurrence. All genes were present in the validation test (log rank, P< 0.004). A multivariable analysis identified GADD45B as an independent marker of disease free survival (Cox regression, P= 0.015). Compared to well-established clinical parameters, as lymph node involvement and extrathyroidal extension, our model presented superior value in predicting recurrence (AUC= 0.741; CI 95%= 0.615-0.868). Conclusion. GADD45B was identified as an independent marker of poor outcome and can be useful to better stratify PTC patients according to the risk of recurrence.

3386F

Targeting the metabolic pathways of cancer is a hot topic for drug discovery. An emerging theme in cancer biology is that a number of lipid metabolic genes, consistently over-expressed in human cancers, are critical to cellular transformation and in maintaining the transformed state. OLR1 gene, encoding the cell membrane receptor LOX-1 (lectin-like oxidized low density lipo-protein receptor), emerged as one of the most important for maintaining the malignant phenotype. LOX-1 may have at least two independent pro-oncogenic mechanisms of action, one based on activation of Nf-kb signaling pathway and the other as regulator of lipogenesis. We observed that LOX-1 is overexpressed in human breast and colorectal carcinomas, at different stages of disease, as compared to normal tissues. Moreover, we demonstrated that the metabolic oncogene FASN (fatty acid synthase), is overexpressed in the breast and colon neoplastic cells. Blocking OLR1 gene in a colon cancer cell line (LDL1), we observed a strong down regulation of FASN and an evident inhibition of the secreted isoform of Clusterin (sCLU) a cytoprotective protein, involved in tumor progression. Consequently, a strong up regulation of the nuclear clusterin isoform (nCLU), was observed. These data suggest a role of LOX-1 in the cancer insurgence and progression. Therefore, this protein could be considered a novel target for cancer therapy.
3387W
Increased frequencies of GSTM1 and GSTT1 null genotypes in Indian patients with Leukemia. S. Caplash1, S. Kaur2, R. Arora2. 1) Human Genetics, Punjabi University, Patiala, Punjab, India; 2) Oshwal Cancer Hospita, Ludhiana.

Glutathione S-transferases (GST) isoenzymes play a significant role in phase II biotransformation and detoxification of many xenobiotics including environmental carcinogens, pollutants and drugs. Both the genetic polymorphisms and expression pattern of GST genes may have a major impact on cancer susceptibility, inter-individual variability in the prognosis, drug effects and toxicity. Of these, GSTT1 and GSTM1 isoenzymes are highly polymorphic with homozygous deletion of either or both genes resulting in absence of enzyme activity. We conducted a case-control study involving 142 Leukemia patients and 60 normal healthy controls from Punjab (North India). Multiplex PCR was carried out to determine GSTM1 and GSTT1 polymorphism. The frequency of individuals carrying GSTM1 and GSTT1 null genotypes was higher among Leukemia patients (42 % and 23 %) as compared to the control group (38 % and 17 %), but the difference found was not statistically significant (p>0.05). Follow up of patients is in process to analyze the association of GST genotypes with differential chemotherapy drug response.

3388T
Two different BRCA2 mutations found in a multigenerational family with a history of breast, prostate, and lung cancers. D. Caporale, E. Swanson. Biology, University of Wisconsin-Stevens Point, Stevens Point, WI.

Breast and lung cancer are two of the most common malignancies in the United States, causing approximately 40,000 and 160,000 deaths each year, respectively. Approximately 5 to 10 percent of breast cancer cases are hereditary, with about 84 percent of those cases due to mutations in two large breast cancer predisposition genes, BRCA1 and BRCA2. These are normally tumor-suppressor genes associated with DNA repair. Since the discovery of these two genes in 1994 and 1995, respectively, several other breast cancer predisposition genes have been identified, including the CHEK2 gene, which is another tumor-suppressor gene. Recently, studies have begun investigating the roles of BRCA1 and BRCA2 in lung cancer. We conducted two case studies, one on an extended family of Italian heritage with several cases of breast cancer and associated cancers through multiple generations, and another on a non-blood relative of Scottish/Irish descent who was consecutively diagnosed with breast and lung cancer. Cancer history and environmental risk factors were recorded for each family member. To investigate possible genetic risks, we screened for mutations in specific hypervariable regions of the BRCA1, BRCA2, and CHEK2 genes. DNA was extracted and isolated from the individuals' hair follicles. PCR, allele-specific PCR (AS-PCR) and DNA sequencing were performed to identify and verify the presence or absence of mutations in these regions. Genotypes of each family member were determined and carriers of mutations were identified. Specifically, three Italian family members were found carriers of the BRCA2-3036del4 mutation, a 4-nucleotide deletion in an exon, which is a nonsense mutation that causes a frameshift in the genetic code, rendering the BRCA2 product nonfunctional. The individual with breast and lung cancer was not a carrier of this mutation, but rather a carrier of the BRCA2-6503delTT, which is also a nonsense mutation but more common in the Irish heritage.

3389F
Negotiating a Minefield: Which Variants to Return in a Large Prospective Whole Exome Sequencing Project? A. Church1,2, E. Van Allen3,4,6, E. Hiller4, I. Rainville3, H. Rana2,4, D. Treacy2, K. Karalis2,5, F. Huang3,4,6, M. Giannakis3,4,6, F. Wilson3,4,6, E. Stover4,6, J. Bohkan1, L. Sholl1, N. Lindeman1, J. Garber3,6, N. Wagle3,4,6, L. Garraway2,4,6. 1) Pathology, Brigham and Women's Hospital, Boston, MA; 2) Pathology, Boston Children's Hospital, Boston, MA; 3) Medical Oncology, Dana Farber Cancer Institute, Boston, MA; 4) Medicine, Dana Farber Cancer Institute, Boston MA; 5) Wyss Institute, Boston, MA; 6) Broad Institute, Boston, MA.

The CanSeq study is a prospective whole exome sequencing project currently enrolling patients at the Dana Farber Cancer Institute and Brigham and Women's Hospital. Whole exome sequencing is performed at the Broad Institute on both somatic cancer tissue as well as germline DNA derived from blood. As a large collaborative effort, the aims of the project include identification of somatic and germline alterations, the identification of potentially actionable and consequential variants, and to prioritize events for treatment. Enrolled patients, who have undertaken informed consent, receive curated results of both somatic and germline sequencing.

Herein we describe the process we have undertaken to identify genes and variants of clinical significance to return to patients and their treating physicians. Lists of genes include somatic (122 genes), germline risk and carrier (149 cancer and 35 non-cancer genes), pharmacogenomics (70 cancer and 18 non-cancer genes). Our process has evolved over years and incorporates available literature, committee guidelines including the recent ACMG recommendations, ethical considerations such as beneficence, nonmaleficence and justice, and the consensus of our expert collaborators.
3391T
MiDNA copy number alteration and association with clinicopathological features in Breast Cancer patient. M. Ghaffarpour, M. Houshmand. Medical Genetics, National Institute of Genetic Engineering and Biotechnology, Tehran, Tehran, Iran.

Large-scale deletion in mitochondrial DNA (mtDNA) has been extensively detected in various human cancers. However, it still remains unclear whether the alterations in mtDNA content are related to the clinicopathological parameters and patient with breast cancer. To explore the role of mtDNA copy number in breast cancer etiology and association between clinicopathological features, in the present study, we performed the copy number of mtDNA in 59 cases of breast tumors and paired non-tumor tissues using quantitative real-time PCR. Our data showed that the level of mtDNA was significantly increased in tumor tissues in compared to the adjacent non-tumor counter-parts (P<0.05). The increased copy number in mtDNA was not significant association with clinicopathological features such as age, grade, stage, lymph node involvement, ER, PR and Her-2/neu expression. However, there was a significant association between decreased overall Survival (Syr%) and increased copy number in mtDNA (p=0.049). Together, our results suggested that high copy number of mtDNA may be involved in breast neoplastic transformation and mtDNA content might be potentially used as a tool to prognosis factor with the other risk factors.

3392F
TNFA-308 G>A is associated with HER-2/neu expression in women with breast cancer. L. Gomez-Flores-Ramos1, A. Escoto-De Dios1, A. Ramos-Silva1,2, R. Ramirez-Patiño1,2, I. Gutierrez-Hurtado1,2, MP. Gallegos-Arreola1. 1) Laboratorio de Genetica Molecular, CIBIO, IMSS, Guadalajara, Mexico; 2) Doctorado en Genetica Humana, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara; 3) Doctorado en Farmacologia, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara.

Introduction. The tumor necrosis factor-alpha gene (TNFA) plays an important role in cell proliferation, differentiation, apoptosis, lipid metabolism, coagulation, insulin resistance, and endothelial function. In vitro and in vivo experiments have shown the tumorigenic effects of TNF, and elevated plasma levels of TNF have been associated with a poor prognosis in Breast Cancer (BC). The TNFA-308G>A (rs1800629) polymorphism has an effect on the gene expression, increasing the production of TNF protein. The aim is to determine the association between the TNFA-308G>A in women with BC and Her-2/neu expression. Methods. Blood samples were collected after a written informed consent approved by the ethical committee from 371 controls and 465 patients with clinical and histological confirmation of BC living in Guadalajara, Mexico. Genotyping was performed by RFLPs. The genotype frequencies at the control group were in Hardy-Weinberg equilibrium. Statistical analyses were performed using PASW Statistic Base 18 software, 2009 (Chicago, IL). Results and discussion. Our results show an association between the genotypes GA-AA and the expression of Her-2/neu in BC cells (OR 1.6, 95% CI 1.06-2.4, p=0.025). The patients with the genotypes GA-AA at tumor stages I-II were also associated with Her-2/neu positivity (OR 2.5, 95% CI 1.31-4.8, p=0.004) as well as patients with BMI >30-40 and Her-2/neu (OR 2.8, 95% CI 1.2-6.6, p = 0.016).BC patients the presence of HER-2/neu has been correlated with metastases, insulin resistance and overexpression of fatty acid synthase, and its expression is regulated by metformin in in vitro models. HER-2/neu serum concentrations in patients with diabetes type 2 were associated with serum TNF receptor 1, thus TNF might be behind comitant insulin resistance and EGF resistance, leading to increased circulating HER-2/neu levels. Although the mechanisms of how the overexpression of HER-2/neu oncogene induces resistance to TNF is unknown; different studies have been observed their overexpression is associated with poor prognosis in BC, because it induces metastatic potential and resistance in cancer cells. Another mechanisms proposed is that the overexpression of Her-2/neu is mediated by signaling pathways and cytotoxicity, suggesting that the overexpression of TNF receptor-1 is important in TNF sensitivity in HER-2/neu-overexpressing cancer cells.

3393W
The anti-cancer activity of propolis in colorectal cancer. M. Gunduz1,2, G. Nas1, M. Acar1, O.F. Hatipoglu2, B. Yilmaz2, G. Kaya1, E. Gunduz1. 1) Medical Genetics, Turgut Ozal University Medical Faculty, Ankara, Turkey; 2) Department of Otologyngology Head and Neck Surgery, Ankara, Turkey.

Colorectal cancer is the third most commonly diagnosed cancer in the world. It originates from the epithelial cells lining the colon or rectum. Colorectal cancer is commonly observed in developed countries and is the second deadliest type of cancer. Colorectal tumorigenesis proceeds through an accumulation of specific molecular alterations and one of the most important of these mechanisms is inhibition of apoptosis. The Inhibitors of Apoptosis (IAP) are a family of functionally and structurally related proteins, which serve as endogenous inhibitors of programmed cell death (apoptosis). Survivin is a member of the IAP family and functions as an inhibitor of caspase activation, thereby leading to negative regulation of apoptosis. This has been shown by disruption of Survivin induction pathways leading to increase in apoptosis and decrease in tumor growth. Propolis is a resinous mixture that honey bees collect from tree buds, sap flows, or other botanical sources. It is used as a sealant for unwanted open spaces in the hive and possesses anti-microbial, anti-oxidative, anti-ulcer and anti-tumor activities. The survivin protein is expressed highly in most human tumors but is completely absent in terminally differentiated cells. Our study investigated the effect of propolis onHT29 colorectal cancer cells. Propolis exhibits anti-cancer activity in a dose dependent manner by inhibition of survivin gene expression. Decreased Survivin expression results in a decrease in tumor growth and prevents carcinogenesis by inducing programmed cell death. Consequently, Propolis may become a strong candidate molecular agent for cancer therapy.

3394T
Identification of Complete Hydatidiform Mole Pregnancy-associated MicroRNAs in Plasma. Y. Hasegawa1, K. Miura1, A. Higashijima1, S. Miura1, J. Tsukamoto1, S. Abe1, K. Kinoshita2, H. Matsuura3, K. Yoshura4, S. Abe1, K. Miyazaki1. 1) Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 2) Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, , Nagasaki, Japan.

Objective: The aim of this study was to identify the complete hydatidiform mole (CHM) pregnancy-associated microRNAs in plasma. Methods: First, by comparative analysis of next generation sequencing-generated microRNAs expression profiles, we selected the candidate CHM pregnancy-associated microRNAs in plasma, which showed a higher expression in CHM tissues than in normal villous tissues, but no expression in blood cells. Then, expression levels of these microRNAs in CHM tissues (n:14) and normal villous tissues (n:20) were investigated to identify CHM-associated microRNAs with increased expressions in CHM tissues than in normal villous tissues. Subsequently, when the plasma concentrations of cell-free CHM-associated microRNAs were significantly higher in CHM pregnant women (n:14) than in uncomplicated pregnant women (n:20), these microRNAs were identified as CHM-associated microRNAs. Conclusively, Propolis may become a strong candidate molecular agent for cancer therapy.
**3395F**

Eukaryotic translation initiation factor 4E (eIF4E) expression is associated with breast cancer tumor phenotype and predicts survival after anthracycline chemotherapy treatment. T. Heikkinen1, T. Korpeila2, R. Fagerholm1, S. Khan1, K. Altmäkki3, P. Heikkilä4, C. Blomqvist5, O. Carpen6, H. Nevanlinna1. 1) Department of Obstetrics and Gynecology, Helsinki university central hospital, Helsinki, Finland; 2) Department of pathology and Medicity laboratory, University of Turku, Turku Finland; 3) Department of medical genetics, Helsinki university central hospital, Helsinki, Finland; 4) Department of pathology, Helsinki university central hospital, Helsinki, Finland; 5) Department of oncology, Helsinki university central hospital, Helsinki, Finland; 6) Turku university hospital, Turku Finland.

Abnormal translation of mRNAs, which frequently occurs during carcinogenesis, is among the mechanisms that affect the expression of proteins involved in tumor development and progression. Eukaryotic initiation factor eIF4E is a key regulator of translation of many cancer-related transcripts, including VEGF, Cyclin D1, and MMP-9. It is over-expressed in many cancers and has been associated with worse survival of cancer patients. The protein levels of eIF4E were determined with immunohistochemistry (IHC) in 1,233 paraffin-embedded breast tumors on tissue microarrays. The effects of the IHC expression level were analyzed on tumor characteristics and patient survival. The survival analyses were also stratified by tumor properties and adjuvant chemotherapy treatment. A number of 1,085 tumors were successfully stained. A high level of eIF4E protein expression was associated with several characteristics of aggressive breast cancer, namely grade, estrogen and progesterone receptor negativity, HER2 receptor positivity and high expression of p53 and Ki67 (p<0.001). High eIF4E expression was also associated with worse breast cancer-specific survival with a hazard ratio (HR) of 1.99 (95% CI:1.32-3.00, p=0.0008), indicating high expression of eIF4E was found to be an independent prognostic factor. High eIF4E expression was also associated with worse survival after detection of distant metastasis (HR=1.88, 95% CI:1.20-2.94, p=0.0060). In the stratified survival analyses the survival effect was strongest among patients treated with anthracycline chemotherapy (HR=3.34, 95% CI:1.72-6.48, p=0.0002), whereas no effect was seen among patients who had not received anthracycline chemotherapy treatment.

**3396W**

Personalized Genomics of Metformin Therapy for Improved Cancer Survival. C. C. Iverson1,2, H. Xu3, Q. Chen4,5, A. Shahi6, Q. Dai6, J. Warrington7,8, M. C. Aldrich1,2,5, L. Olson9, D. C. Crawford1, D. M. Roden2,7, J. C. Denny1,2,5, M. C. Aldrich1,2,5. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Thoracic Surgery, Vanderbilt University, Nashville, TN; 3) The University of Texas School of Biomedical Sciences at Houston, Houston, TX; 4) Department of Biostatistics, Vanderbilt University, Nashville, TN; 5) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 6) Division of Epidemiology, Vanderbilt University, Nashville, TN; 7) Department of Medicine, Vanderbilt University, Nashville, TN; 8) Department of Pharmacology, Vanderbilt University, Nashville, TN.

Increasing evidence supports a role for metformin as a therapeutic for improved cancer survival. To determine the potential use of metformin in a personalized genomic medicine approach for treating breast cancer, we performed a genome-wide association study of overall survival in cancer patients. Using the Vanderbilt University large-scale bioRepository linked to electronic medical records, we identified 1,106 individuals (674 with type 2 diabetes (T2D), 432 non-diabetics) of European ancestry diagnosed with cancer between 1995 and 2010. Among cancer patients with T2D, 461 individuals were on metformin therapy and 213 were treated with other T2D medications. Patients were followed a median of 2.3 years (IQR: 1.3-3.8) for overall mortality or loss to follow-up. Participants were genotyped using either the HumanOmniQuad or Illumina Omni-Quad genotyping platforms. We estimated hazard ratios (HRs) and accompanying 95% confidence intervals (CIs) for overall mortality using Cox proportional hazard models, adjusted for age and sex, to evaluate associations between genetic variants and mortality. In an additive model, we used single nucleotide polymorphisms to correct for potential population stratification occurring in the study. Among cancer patients who were metformin users, we identified one marker, rs153047, near genome-wide significant and associated with a two-fold higher mortality (HR=2.22, 95% CI: 1.91-2.54). This SNP was associated with mortality in T2D cancer patients on other drugs or non-diabetic cancer patients (p=0.20 for all groups). Assessment of gene-environment interactions demonstrated that the association with rs153047 may vary by metformin use (p=0.15). Among individuals homozygous for the reference allele, metformin use was associated with a significantly reduced mortality compared to non-metformin use (HR=0.58, 95% CI: 0.25-0.92, p=0.0023). Metformin use was not significantly associated with improved survival for individuals with the variant allele (HR=0.98, 95% CI: 0.47-1.94, p=0.95). This marker is located within SNTB2 on chromosome 16q22.1, which has been previously shown to be associated with invasive breast cancer. Additional SNPs on chromosomes 3, 10, 12, 13, and 16 had p<0.02 and deserve further investigation. Our findings suggest the survival benefit of metformin therapy may be modulated by specific genomic loci. If replicated, this finding may be useful to target metformin therapy in cancer patients.

**3397T**

Comparison of mRNA expression profiles in familial and sporadic breast cancers in Finnish population. S. Khan1, P. Heikkilä2, K. Altmäkki3, C. Blomqvist4, H. Nevanlinna1. 1) Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 2) Department of Pathology, University of Turku and Helsinki University Central Hospital, Helsinki, Finland; 3) Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 4) Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland.

Although majority of breast cancers initiate from somatic alterations, familial aggregation and twin studies have suggested that hereditary predisposing factors are involved in up to one third of all breast cancers. The underlying inherited mutations in breast cancer families may also affect the progression and mRNA expression profiles of the tumors. In order to investigate the differences in mRNA expression profiles of familial and sporadic tumors, we performed a comparative study of mRNA levels in 211 familial and sporadic tumors from Southern Finland. Altogether we had mRNA from 139 sporadic and 72 familial samples. The expression data was quantile normalized and the comparison in mRNA expression levels between familial and sporadic samples were performed by empirical Bayes t-test. By comparing the gene expression in familial vs. sporadic breast tumors overall we did not find any genes with expression greater than 1.2-fold above or below expression threshold. Since the estrogen receptor (ER) status has been shown to affect gene expression profiles in breast cancer, we further analysed the tumor samples according to the ER status. In the ER positive subgroup, we found 3 genes upregulated and 2 genes downregulated above and below 1.2-fold expression threshold in the familial vs. sporadic subgroups. The largest differences between the familial and sporadic subgroup were obtained in ER negative subgroup whereas we identified 25 up- and 25 downregulated genes (9 up- and 6 downregulated genes with greater than 2-fold expression threshold). The results help pinpointing genes with aberrant expression in putative germline mutation carriers compared to sporadic cases, especially in the ER defined subgroups and thus might indicate novel susceptibility genes.
3398F

Several recent studies have involved assaying the sensitivity of a library of cancer cell lines to an array of anti-cancer compounds, in particular the Cancer Cell Line Encyclopedia (CCLE), Genomics of Drug Sensitivity in Cancer (GDS), and the Heiser dataset on which the DREAM challenge was based. Along with growth curve measurements, molecular characteristics of the cell lines are assayed: CCLE for example includes gene expression microarrays (GE), copy number variation (CNV), and oncogene mutation status assays. We extend previous per drug regression analysis using biologically informed dimensionality reduction and performing joint analysis across drugs in order to improve statistical power. We perform two stages of dimensionality reduction on the cell line characteristics, both using biological knowledge. In the first stage, we use a one dimensional factor analysis model for each gene, so that each data modality is considered an observed variable, explained by the latent factor which we envisage as the activation level of this gene. This is distinct from simply averaging across the different modalities: for example, methylation might be negatively correlated with gene expression, which we can account for. In the second stage, we use a hand curated collection of 1967 known pathways, collected from resources including GO, KEGG, and published GWAS hits, to construct a per pathway activation level, again using a one dimensional factor analysis. This approach gives easily interpretable results as the predictive pathways are associated with particular functions. While cancer is an incredibly heterogeneous disease there are characteristics shared across many cancers: disruption of apoptosis, cell cycle regulation, or DNA repair mechanisms, or addiction to specific oncogene pathways. We develop a model that hypothesizes a finite set of such unobserved, latent characteristics, which explain the observed drug sensitivity patterns and which are encouraged to be related to the pathway activation levels found by the dimensionality reduction procedure. The presence or absence of these characteristics confers sensitivity or resistance to specific therapeutic compounds. Our initial results on the cCLE dataset suggest five such latent characteristics are key predictors of drug sensitivity across cell lines and drugs. The top two most frequently used pathways involve vasculature development and TGF-beta signaling.

3399W
Prognostic significance of syndecan-1 expression in colorectal carcinoma. S.H Lee1, E.J Cho2, J.A Yoon3, E.S Jung2, S.Y Kim1. 1) Department of Hospital Pathology, College of Medicine, Seoul St. Mary’s Hospital, The Catholic University of Korea, Seoul, South Korea; 2) Department of Pathology, College of Medicine, The Catholic University of Korea, Seoul, South Korea.

Background: Syndecan-1 (SDC1) is reported to modulate several key processes of tumorigenesis and have variable expression in many cancers. The cause provoking altered expression is not known to date. In this study, we compared SDC1 status with various clinicopathologic parameters and molecular markers to evaluate clinical implications in colorectal carcinoma. Methods: With 219 surgical specimens of colorectal carcinoma, treated between 2008 and 2010 at the Seoul St.Mary’s hospital, the catholic university of Korea. The study group consisted of 136 men and 83 women. The mean age was 62.3 years (range 32-93 years). We screened SDC1 expression using immunohistochemistry and analyzed the relationship between SDC1 expression and various clinicopathological parameters and molecular markers. Results: The tumors were located mainly in the left colon (68.5%) and especially in the rectum (37.9%). There were 206 (94.1%) adenocarcinomas, 10 (4.6%) mucinous carcinomas, and 3 other tumors. Most of the carcinomas were pT3 and pT4 (n=150 and 49, respectively). Regional lymph nodes contained metastases in 138 patients. SDC1 expression was found in cancer cells in 212 cases (96.8%) of colon cancer. Of the SDC1 expression cases, 131 cases dominantly showed membranous immunopositivity, 81 cases showing cytoplasmatic pattern. 154 cases showed mixed membranous and cytoplasmatic pattern. In 3 cases,stromal SDC1 reactivity was noted. It was significantly associated with primary tumor (T) (p<0.001) and peritoneal invasion (p=0.033). SDC1 expression was correlated with EGFR and HR status (p=0.0004). On multivariate analysis, it was not significantly correlated with lymph node metastasis (N), distance metastasis (M), lymphatic and vascular invasion, and K-ras mutation status Conclusions: We showed that SDC1 expression is associated with T stage, status of cancer invasion and EGFR immunohistochemical expression. The expression of SDC1 could be a biomarker of colorectal carcinoma. Further studies are necessary to better understand the role of SDC1 in the progression and invasiveness of colorectal carcinoma.

3400T
Mutations and copy number changes identified in primary brain tumors using complementary analyses and formalin-fixed, paraffin-embedded (FFPE) tissue. A. Ligon1,2,3, P. Ramkisson1,2,3, P. Wen1,2, D. Reardon1,2, E. Lee1,2, M. Rinne1,2, A. Norden2, L. Nayak3, S. Nuland3, L. Doherty4, D. Lafarangi4, L. Brown4, N. Avold4, A. Sanganatagad5, I. Dunn5, N. Lindeman5,1, L. MacConaill6,2, B. Rollins2,3, R. Baroukhin2,4, L. K. Kellison6,1. 1) Department of Pathology, Brigham & Women’s Hospital, Boston, MA; 2) Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Broad Institute, Cambridge, MA.

Routine and integrated genomic approaches for genotyping clinical oncologic specimens have not yet been well-established. Whole genome testing in oncologic specimens can inform diagnosis, prognosis and genotypic stratification with respect to clinical trials. However, the accessibility of such data is constrained by the frequent need to rely on routine formalin-fixed, paraffin-embedded (FFPE) samples as starting material. In this study, whole genome copy number analysis was performed as a routine clinical test for 400 primary adult and pediatric FFPE brain tumors. The tumors, which included meningiomas and gliomas, were analyzed by array-based comparative genomic hybridization (aCGH) using stock 1x1M Agilent SurePrint G3 arrays. Twenty-three copy number aberrations relevant to diagnosis, prognosis or clinical management were scored for each tumor. In addition, as part of a clinical data research program, we profiled somatic mutations using the Oncomap v4.4 platform (Sequenom), which queried 471 known cancer-related mutations in 41 genes. We identified PIK3CA mutations in grade I meningothelial subtype tumors and found that sequence mutations, including IDH1 and IDH2 mutations, were present in ~30% of gliomas, mostly in tumors of WHO grade II or III. Both mutation and copy number data were obtained for a subset of the gliomas, 68% of which were classified as glioblastoma (GBM). Copy number changes identified in GBM varied with the patient age and IDH1 mutation status, and included polyosity 7, EGFR amplification, identification of the EGFRvIII structural variant, as well as losses of CDKN2A, PTEN, RB1 and TP53. Copy number profiles for IDH1(p.R132H) mutant tumors generally showed less complexity than that of tumors with wild-type IDH1. 1p/19q co-deletions were readily identified in IDH1 mutant glioblastomas, but overlaps were not identified with the BRAF duplication event observed in pilocytic astrocytomas and related tumors. In summary, we developed complementary assays for detecting common cancer-specific mutations and genome wide copy number changes from FFPE tissue. Together, these assays have been integrated to generate data that are diagnostic (e.g., 1p/19q co-deletion), prognostic (e.g., IDH1 and IDH2 mutation status), or that illuminate genomic aberrations (e.g., identification of EGFRvIII variant) for use in advancing clinical trial enrollment and therapeutic management.

3401F
A leukemic stem cell score associated with patient prognosis and tumor histopathology in ovarian cancer. B.A. Logsdon1, S.L. Battle2, M.H. Rendi3, R.D. Hawkins2, S.H. Lee2,3. 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Obstetrics, University of Washington, Seattle, WA; 3) Department of Obstetrics and Gynecology, University of Washington, Seattle, WA.

We derive a novel leukemic stem cell (LSC) score based on the gene expression levels of ten genes associated with both patient prognosis and leukemic stem cells in acute myelogenous leukemia (AML). Using gene weights estimated from AML patients across multiple studies, we generate the LSC score in three independent gene expression data-sets for patients with ovarian cancer. The LSC score is significantly associated with poor patient prognosis in ovarian cancer (Cox proportional hazard inverse normal p-value: 8.55x10^-1). The LSC score is also significantly associated with the optimal debulking status of tumors (Fisher meta-analysis p-value: 4.93x10^-3). In data available from The Cancer Genome Atlas on patients with ovarian serous carcinoma we show that the LSC score is associated with abnormal tumor stroma, increased desmoplastic response, and evidence of increased abnormal angiogenesis in patients with high LSC scores, suggesting an opportunity for targeted interventions with anti-vascular endothelial growth factor therapies. This strategy is further motivated by recent efforts to target tumor vascularization with anti-angiogenic therapies in AML. Our results suggest the LSC score is a novel biomarker that is informative for cancer stem cell populations, abnormal tumor histopathology, and patient prognosis in ovarian cancer.
3402W
Coupled Use of Family-based Exome Sequencing and TCGA Germline Data Analysis to Identify Novel Breast and Ovarian Cancer Susceptibility Genes. J.A. Martignetti1, P. Dottino1, M. Babcock1, L. Jara1, K. Moyisch1, L. Sucheston2, S. Lele3, K. Ondus2, S. Sangha3; 1) Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 2) School of Medicine and Biomedical Sciences State University of New York at Buffalo, Buffalo, NY, USA; 3) Human Genetics Program, University of Chile School of Medicine, Santiago, Chile; 4) Station X, Inc., San Francisco, CA, USA.

Family history is the strongest single predictor of a woman’s risk for developing breast and ovarian cancers. Genetic studies seeking to identify hereditary breast and ovarian cancer susceptibility genes have therefore focused on those families with a high incidence of cancer across multiple generations. While they represent the strongest known genetic predictors, BRCA1/2 mutations account for less than half of all families containing two or more breast and ovarian cancer cases in first-degree relatives and explain less than half of the excess familial cancer risk. Our hypothesis is that the residual risk of familial breast and ovarian cancers is attributable to mutations and/or genetic variants in other genes besides BRCA1/2. To identify novel susceptibility genes, we exome sequenced the germlines of a discovery cohort of selected families with hereditary breast and ovarian cancer but who lacked deleterious BRCA mutations. The discovery cohort was sequenced using Illumina technology at the Icahn School of Medicine at Mount Sinai and Beijing Genomics Institute. Permutations of BWA sequence alignment, GATK re-calibration/re-alignment, SRMA re-alignment, GATK variant calling, and VarScan2 variant calling were used to identify small genomic variations (SNVs, small INDELS). These variants were annotated, visualized, analyzed, and evaluated within GenePool™ (Station X, Inc., San Francisco, CA, USA), a discovery database for rare or rare variants with an autosomal dominant model of inheritance. All exome-defined candidate mutations were then triaged by leveraging biological annotations, filtering capabilities, and cross-referencing the interactive genome browser within GenePool™. Once this highest tier of candidates was established, they were then compared within and between families, and candidate genes mutated in multiple families were identified. To further refine this list and begin to gain an understanding of the extent of their mutation in a larger cohort, we compared our findings to germline and somatic whole exome sequence variant frequencies derived from the most up-to-date TCGA breast and ovarian cancer genome data. The somatic and germline variants in the TCGA cohorts were called using bioinformatics pipelines consistent with those used for the discovery cohort. This approach allowed us to highlight the major differences between the two population granularities along with the relevant clinical metadata per sample. The final variant list of novel breast and ovarian susceptibility genes, and functional validation studies will be discussed.

3403T
KISS1 suppressor metastasis gene rs12998 and rs5780218 polymorphisms in Mexican patients with breast cancer. M.C. Moran Moguel1, S.E. Flores Martínez1, L.A. Juarez Aguilar1, J. Sanchez Corona1, I.P. Devalos Rodriguez1, R.C. Rosales2, S.A. Rosales2, G. Guadalajara. Guadalajara, Jalisco, Mexico; 1) Doctorado en Genética Humana. CUCS. Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Doctorado en Genética Humana. CUCS. Universidad de Guadalajara, Guadalajara, Jalisco, Mexico.

Aims. KISS1 is a suppressor metastasis gene that has been associated with inhibition of cellular chemotaxis and invasion attenuating the metastases in melanoma and breast cancer cell lines. Along the KISS1-1 gene have been described at least 130 SNPs however the association of these polymorphisms as genetic markers for metastasis in breast cancer studies has not been investigated. Here we describe two simple PCR-RFLP protocols to identify the 9DelT (rs5780218) and E20K (rs12998) KISS1 polymorphisms in Mexican populations and the allelic, genotypic and haplotype frequencies in Mexican general population (GP) and patients with benign breast disease (BBD) or breast cancer (BC) in any histological stage. Results. Individually, none of the polymorphisms were associated with breast cancer. Since we compared the breast cancer group vs BBD and GP groups, no differences were found (p>0.05); however, for polymorphism rs 12998 comparing GP vs case (BC and BBD) groups, statistically significant differences were found. The haplotype consisting of the wild-type allele of rs12998 and variant allele of rs5780218 polymorphisms (G/del) occurred more frequently in the BC group (0.4266) whereas haplotype G/T (both wild type alleles) was the most prevalent in BBD group (0.4674). Conclusions. Our data indicated that these polymorphisms individually do not confer susceptibility to breast cancer metastasis in our population. However, a possible role as genetic marker in breast cancer metastasis for H1 haplotype (W/Variant) in KISS1 gene, must be analyzed in other populations.

3404F
Initial viral load in cases of single human papillomavirus 16 or 52 persistent infection is associated with progression of later cytological findings in the uterine cervix. K. Owusu1, K. Miura1, S. Abe1, A. Kinoshita1, S. Miura1, D. Hamaguchi1, Kl. Yoshiura2, H. Masuzaki1; 1) Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Nagasaki, Japan; 2) Human Genetics, Graduate School of Biomedical Sciences, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan.

Objectives: The aim of this study was to investigate the relationship between viral load in single human papillomavirus (HPV) 16 or 52 persistent infection and the progression of later cytological findings in the uterine cervix.

Methods: Cervical cytological tests and HPV genotyping tests were repeated within 3-6 months in 305 women with oncogenic HPV. Then, 24 cases of a single HPV52 persistent infection and 24 cases of a single HPV16 persistent infection were identified. Cases of later cytological findings showing progression were defined as ‘progression’ group, while cases of no-change or regression were as ‘non-progression’ group. Relative HPV DNA loads were determined by quantitative real-time polymerase chain reaction and expressed relative to human albumin (ALB) DNA. Differences between the two groups were evaluated. Results: The median relative HPV 52 DNA load was 2.211 in the progression group and 0.022 in the non-progression group (Mann-Whitney U test, P=0.003). The median relative HPV 16 DNA load was 4.206 in the progression group and 0.103 in the non-progression group (P=0.001). Conclusions: HPV 52 or 16 DNA load, which is measured by quantitative real-time methods, may be used as a short-term marker to identify women at high risk for progression of cervical cytological findings.

3405W
Conventional and Molecular Cyto genetic Techniques in Comparison with mRNA and DNA Based Quantitative Real-Time PCR to Monitor Minimal Residual Disease in Chronic Myeloid Leukemia. I.S. Pagani1, C. Pirrone2, D. Pigni3, O. Spinelli2, C. Boroni3, T. Intermesoli1, U. Giussani3, F. Pasquali4, L. Lo Curto1, A. Lanfranchi3, F. Porta3, A. Rambaldi4, G. Porta5; 1) Department of Clinical and Experimental Medicine, Insubria University, Varese, Italy; 2) Paolo Belli Haematology Division, Papa Giovanni XVIII Hospital, Bergamo, Italy; 3) Hematology and BMT Unit Children’s Hospital Spedali Civili Brescia, Italy; 4) Dept. of Experimental Medicine and Surgery, University of Rome Tor Vergata, Italy.

The Imatinib mesylate (IM) is the first line therapy against Chronic Myeloid Leukemia (CML), through the inhibition of the BCR-ABL1 proliferation pathway. Despite its efficacy in prolonging overall survival, discontinuation of treatment is associated with molecular relapse. So, to maintain an operational cure IM is required indefinitely, despite financial cost to the community and side effects. To evaluate the degree of response to therapy and to highlight the persistence of the disease after treatment, patient should be monitored routinely. The gold standard for diagnosing CML is the cytogenetic analysis, a direct not-sensitive method to detect Ph-positive cells. Quantitative real-time RT-PCR (qRT-PCR) based on mRNA is the most sensitive technique available for the detection of BCR-ABL1 transcripts and it is used to follow the progression of CML after initial diagnosis and treatment. However mRNA levels are not directly related to number of leukemic cells and negative results are difficult to interpret, because undetectable levels of chimeric transcript can reflect either an effective elimination of leukemic cells, or the presence of a leukemic cell transcriptionally silent. Here we will propose a new sensitive approach to directly detect the number of leukemic cells using a DNA-based biomarker specific for each patient. For the first time we developed a DNA Q-PCR assay based on the genomic breakpoint found with Next-generation sequencing and a formula to calculate the number of Ph+ cells. We monitored eight CML patients in their early chronic phase and in follow-ups up to 8 years under Imatinib treatment. We carried out patient specific Q-PCR assays to monitor minimal residual disease, testing the same samples in parallel by cytogenetic analysis and by standard qRT-PCR. In all positive samples for chimeric transcript we measured positive levels of corresponding genomic DNA, confirming the sensitivity of the method. Furthermore as a novelty, we showed the persistence of leukemic cells transcriptionally-silent by Q-PCR in 33% of samples with undetectable levels of mRNA. Finally, we applied our technique in the evaluation of BCR-ABL1 presence in CD34+ sorted cells in order to enrich the cancer stem cells. In conclusion the DNA Q-PCR is a sensitive and direct technique to identify leukemic cells and patients that could be candidate to stop the therapy. Thanks to AIRC, AIL.
3406T

ALK gene copy number gain in a series of 350 NSCLC: a single institution experience. L. Pecciarini1, A. Talancó1, G. Grassini4, E. Del Cin1, S. Foresti1, C. Lazzari1, V. Gregoretti2, M.G. Cangi1, C. Dogliotti4.
1) Pathology Unit, San Raffaele Scientific Institute, Milan, Italy; 2) Oncology Department, San Raffaele Scientific Institute, Milan, Italy.

Anaplastic lymphoma kinase (ALK) gene translocations are involved in the tumorigenesis of a small group of non-small cell lung carcinomas (NSCLCs) and identify patients sensitive to ALK inhibitors. Therefore fluorescence in situ hybridization (FISH) analysis for ALK rearrangements is routinely performed in NSCLC patients for correct therapy selection. ALK gene copy number amplification (GCN) gains are known to play an oncogenic role in tumors such as neuroblastoma, but they are poorly characterized in NSCLC. Indeed in our experience ALK GCN changes represent common incidental findings of the ALK FISH test in NSCLCs. In this study we report the prevalence of ALK GCN gain and its correlation to ALK protein expression, epidermal growth factor receptor (EGFR) gene and v-ki-ras2 Kirsten sarcoma viral oncogene homolog (KRAS) gene mutational status, and clinicopathological data in a series of 350 NSCLC patients treated at our Institution. ALK translocations and gene status was evaluated by fluorescence in situ hybridization (FISH), using a commercial split-apart ALK (2p23) probe. Specimens positive for a ALK split signal were studied for echinoderm microtubule-associated protein-like 4 (EML4, 2p11) gene involvement using a three color FISH probe. Gain of ALK GCN was defined in 2 categories: low gain = 4<ALK<12 copies in ≥40% of tumor cells and high gain = ALK ≥12 copies in ≥25% of tumor cells. ALK expression was assessed by immunohistochemistry. EGFR and KRAS mutational status were also assessed by both direct sequencing and pyrosequencing. ALK translocations were identified in 10% of the analyzed cases and none of them were either EGFR or KRAS mutated. Interestingly ALK GCN gains were observed in 15% of the ALK translocation negative cases: 10% with low gain and 5% with high gain. EGFR gene and KRAS gene were mutated in respectively 20% and 8% of the samples negative: 10% with low gain and 5% with high gain. EGFR gene and KRAS gene mutational status were also assessed by both direct sequencing and pyrosequencing. ALK translocations, but not ALK GCN gains were associated with ALK protein overexpression. In order to better characterize ALK GCN gains in NSCLCs, correlation with clinicopathological data and both EGFR gene and KRAS gene mutations are under evaluation and results will be presented.

3407F

Shortened telomere length and survival in bladder cancer patients. A. Russo1,2, F. Modica1, S. Guarnera1, G. Fiorito1,2, A. Allione1, B. Pardini1, R. Cotelli1,2, F. Ricciarini1, A. Bosio1, G. Casetta1, G. Cucchiare1, P. Dastelletti1,2, P. Galletto1, L. Rolle1, A. Zitella1, D. Fontana1, P. Vinesi1,2, C. Sacerdote6, G. Martelli2,1, 1) Human Genetics Foundation, Turin, Turin, Italy; 2) Department of Medical Sciences, University of Turin, Turin, Italy; 3) Department of Urology, University of Turin, Turin, Italy; 4) Department of Urology 2, San Giovanni Battista Hospital and University of Turin, Turin, Italy.

Telomeres are the specialized DNA-protein structures composed of a variable number of simple repetitive nucleotide repeats (TTAGGG) that cap the ends of linear eukaryotic chromosomes. They are responsible for the protection of the chromosome ends from nucleolytic degradation, end-to-end fusion, irregular recombination, and other lethal events to a cell. Replicating somatic cells in vivo lose ~50-105 bp per year, because of DNA polymerase’s inefficiency and DNA end-branch migration. Telomere shortening results in critically short telomeres, inducing cellular senescence and apoptosis. If protective mechanisms, such as the TP53 tumor-suppresser gene, are inactive, thus allowing continued proliferation, telomeres become extremely short and dysfunctional; end-to-end fusions ultimately cause chromosomal instability that may contribute to the development of several types of cancer. Moreover, as a marker of cumulative cellular aging, short telomere length may also be associated with increased risk of early death in other individuals. In the present study, we analyzed blood telomere length (TL) in both cancer patients and normal people for up to 18 years. We found a significant correlation between telomere length and number of deaths, and that shortened telomeres act as an independent prognostic predictor for survival of bladder cancer patients.

3408W

Generating Hypotheses for Targeted Therapies with Cancer In silico Drug Discovery Tools. F.A. San Lucas1,2, J. Fowler2, S. Kopetz3,4, E. Vitale1,2, P. Scheer1,2.
1) The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, TX; 2) Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX; 3) Clinical Cancer Prevention, The University of Texas M.D. Anderson Cancer Center, Houston, TX; 4) GI Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX.

We have developed Cancer In silico Drug Discovery (CIDD) Tools to empower clinical researchers to perform preliminary analyses for drug discovery projects by facilitating the setup, execution and evaluation of in silico cancer drug experiments. CIDD is a toolset that integrates genomic mutation data, tumor gene expression profiles and drug-induced gene expression profiles from public data sets to help researchers generate hypotheses for three general problems: (1) to determine if a particular somatic mutation or a set of mutations in a cancer is functional, producing unique gene expression signatures, (2) to find candidate drugs to treat, or repress, these alterations in gene expression, and (3) to identify cell-lines for subsequent lab experimentation that most closely represent the cancer being studied. CIDD integrates publicly available experimental data and annotation databases, such as data from The Cancer Genome Atlas (TCGA), the Connectivity Map (CMap) and the Cancer Cell-line Encyclopedia (CCLE) to perform the in silico cancer experiments. CIDD characterizes generated tumor expression signatures using MSigDB and describes candidate drugs by integrating data from annotation databases such as DrugBank and SuperDrug. An end result of a CIDD execution is a statistically derived, biologically interpretable candidate drug list. This empowers clinical researchers to make biologically informed decisions on candidate drugs. To illustrate its use, we applied CIDD to the study of BRAF-mutant colorectal cancers. CIDD identified BRAF-mutant (V600E) samples from the TCGA colorectal cancer (CRC) project, extracted RNA-seq and microarray gene expression data for BRAF-mutant and BRAF-wildtype samples, and identified candidate drugs targeting BRAF-mutant colorectal cancers. CIDD identified candidate drugs for 7 possible cell-lines from the CCLE as the best candidates on which to test the drugs. CIDD is written in Python and requires R. CIDD will be available at http://scheer.org/software.

3409T

1) Biochemistry, PGIMER, Chandigarh, Chandigarh, India; 2) Urology, PGIMER, Chandigarh, Chandigarh, India; 3) Histopathology, PGIMER, Chandigarh, Chandigarh, India.

The present study was conducted to explore the expression and localization of Liver/Bone/kidney alkaline phosphatase (L/B/K ALP) on renal tubular cells in clear cell renal cell carcinoma (RCC). A total of 50 patients of histopathologically confirmed cases of RCC were included in this study. The L/B/K ALP protein level was determined by immunohistochemistry, immunofluorescence and flow cytometry in renal cell carcinoma and adjacent normal renal parenchyma tissue. The mRNA expression of L/B/K ALP was detected using real time PCR. All the parameters were statistically analyzed. Immunohistochemistry showed mean immunoreactivity in all normal renal sections (n = 50) was 2.9. Whereas, mean L/B/K ALP immunoreactivity in all tumor samples (n = 50) was 0.7. Similarly, immunofluorescence showed mean L/B/K ALP immunoreactivity in all normal renal sections (n = 50) was 3.2. Whereas, it was 0.5 in tumor (n=50). The flow cytometric studies documented a significant reduction in the ALP presentation on microvillus membrane of RCC (291.9±16.8 vs 191±15.9; P<0.05). Further, Real time PCR analysis revealed a significant upregulation in the ALP mRNA transcript in the RCC tissue (0.65±0.09 vs 1.0±0.18; P<0.01). These findings conclude that reduced activity of ALP in BBM could be associated with reduced growth of RCC cells. CIDD is written in Python and requires R. CIDD will be available at http://scheer.org/software.

3409T
Hif1a is a transcription factor regulating the expression of multiple genes responsible for the cancer development and metastasis. The expression of Hif1a gene is regulated by transcriptional and post-transcriptional mechanism which is related with the proteosomal degradation of Hif1a protein by PHD family proteins. It has been reported that Hif1a protein itself also regulates the expression of PHD family genes by a feedback mechanism. The high level of Hif1a expression was found in various types of cancer but the expression levels of PHD family genes were remain unclear. Recent studies on hematologic malignancies indicated the differential expression of Hif1a gene in leukemia cells. In particular, chronic myeloid leukemia (CML) stem cells express Hif1a gene accompanied by increased survival capacity. But still there is no data about the expression level of the PHD genes in leukemia cells. Our study was performed to investigate the expression of Hif1a and PHD genes in CML cells. The expression levels of Hif1a and PHD genes compared among each other and the relationship between the BCR-ABL which is the marker gene of CML, and Hif1a expression analyzed. Peripheral blood and bone marrow samples of 29 CML patients, who have been receiving imatinib therapy, were obtained to investigate mRNA expression of Hif1a, PHD1 and PHD2 genes. RPL13A gene was used as a reference gene. Real time quantitative PCR and REST analysis were performed to measure expression levels. Hif1a and PHD2 gene expressions were higher in BCR-ABL negative CML patients compared to BCR-ABL positive patients. There was no difference in the expression of PHD1 gene among these two groups. The patients who have a high level of Hif1a expression also showed high level of PHD2 expression but PHD1 expression levels were unchanged. These results indicate that Hif1a and PHD2 genes have a similar expression pattern among CML patients. According to our data, the expression of PHD1 and PHD2 genes is most likely regulated by a common mechanism responsible for the regulation of both PHD1 and PHD2 genes. The relationship between Hif1a and PHD family genes needs to be further analyzed in CML.

Applicability of high-resolution multicapillary electrophoresis for molecular characterization of immune gene rearrangement profiles in acute lymphoblastic leukemia (ALL). H. Trautmann, O. Aitlio Clark, O. Salmi, A. Timuragaoğlu, M. Okur, O. Yuçel, M. Uluabas, K. Eker Güler, N. Sayin Ekinci, I. Karađoš, L. Undar. 1) Department of Medical Genetics, Akdeniz University Medical Faculty, ANTALYA, Turkey; 2) Department of Hematology, Akdeniz University Medical Faculty, ANTALYA, Turkey; 3) Laboratory of Hematology, Akdeniz University Hospital, ANTALYA, Turkey.

Hif1a is a transcription factor regulating the expression of multiple genes responsible for the cancer development and metastasis. The expression of Hif1a gene is regulated by transcriptional and post-transcriptional mechanism which is related with the proteosomal degradation of Hif1a protein by PHD family proteins. It has been reported that Hif1a protein itself also regulates the expression of PHD family genes by a feedback mechanism. The high level of Hif1a expression was found in various types of cancer but the expression levels of PHD family genes were remain unclear. Recent studies on hematologic malignancies indicated the differential expression of Hif1a gene in leukemia cells. In particular, chronic myeloid leukemia (CML) stem cells express Hif1a gene accompanied by increased survival capacity. But still there is no data about the expression level of the PHD genes in leukemia cells. Our study was performed to investigate the expression of Hif1a and PHD genes in CML cells. The expression levels of Hif1a and PHD genes compared among each other and the relationship between the BCR-ABL which is the marker gene of CML, and Hif1a expression analyzed. Peripheral blood and bone marrow samples of 29 CML patients, who have been receiving imatinib therapy, were obtained to investigate mRNA expression of Hif1a, PHD1 and PHD2 genes. RPL13A gene was used as a reference gene. Real time quantitative PCR and REST analysis were performed to measure expression levels. Hif1a and PHD2 gene expressions were higher in BCR-ABL negative CML patients compared to BCR-ABL positive patients. There was no difference in the expression of PHD1 gene among these two groups. The patients who have a high level of Hif1a expression also showed high level of PHD2 expression but PHD1 expression levels were unchanged. These results indicate that Hif1a and PHD2 genes have a similar expression pattern among CML patients. According to our data, the expression of PHD1 and PHD2 genes is most likely regulated by a common mechanism responsible for the regulation of both PHD1 and PHD2 genes. The relationship between Hif1a and PHD family genes needs to be further analyzed in CML.

Investigation of Resveratrol AT-101’s Molecular Targets of Apoptotic Effect to the Hormon Sensitive and Hormon Nonsensitive Prostate Cancer Cell Lines: YUKSEL YEN, N. AKTEPE, C. KISIM, H. ATMACA, B. KARACA. 1) Department of Medical Biology, Medicine Faculty of Ankara University, Ankara, Turkey; 2) Department of Nursing, Health High School of Artuklu University, Mardin, Turkey; 3) Department of Molecular Biology and Genetics, Faculty of Arts and Sciences- Celal Bayar University, Manisa, Turkey; 4) Department of Oncology, Medicine Faculty of Ege University, Izmir, Turkey.

Objective: Resveratrol, which is used in the cancer treatment, shown to have acquired DNA damages and apoptosis-potentiating action by including antiangiogenic activity, is a phytoalexin derived from the skin grapes. AT-101, an (-) enantiomer of gossypol, is a potent anticancer agent that it was reported to be an inhibitor of Bcl-2/Bcl-XL in the study, the possible synergistic cytotoxic and apoptotic effects of Resveratrol in combination with AT-101 were investigated in human hormone refractory prostate cancer cell lines, PC-3, DU-145 and LNCaP in a time- and dose-dependent manner.

Material and Methods: Cytotoxicity was determined by XTT Cell Proliferation Kit (Roche). Drug synergy was assessed by using CalcuSyn 2.1 software (Biosoft). Apoptosis was detected by Cell Death Detection Elisa Plus Kit (Roche) and confirmed by Caspase-Glo 3/7 Assay. The expression levels of PHD proteins were assessed by human apoptosis antibody array. Expression levels of proteins associated with apoptosis were investigated by Human Apoptosis Array Kit (R&D Systems, UK).

Results: The IC50 values of AT-101 and Resveratrol were found 5.51 µM, 10 µM, 7.5 µM and 114 µM, 166 µM, 150.5 µM in human prostate cancer cell line DU-145, 3410F, 630, respectively. The combination treatment (100 µM + 5 µM AT-101) was shown to have strong synergistic cytotoxic and apoptotic effects in DU-145 cell line, at 72 hours. The combined use of AT-101 and Resveratrol resulted in an increase in Bax, Fas/TNFSF6, Pro-Caspase-3, Cleaved-Caspase-3, FADD, Phospho p53 (S46), p53, p21/CIP1/CDKN1A, Cytochrome c, Smac/DIABLO and a decrease in Bcl-2, Bcl-xL, XIAP, p27/Kip1, cIAP-1 proteins in DU-145 prostate cancer cells.

Conclusion: Our data revealed that the combination of AT-101 with Resveratrol may hold great promise for development as a novel chemotherapeutic approach to human prostate cancer. In determined synergism combination rates, it needs confirmation of promising cytotoxic and apoptotic effects in other prostate cancer cell lines; and also clarification of impact mechanism and confirmation of these data in animal experiments.
3414W
Intrafocal Heterogeneity of Gene Rearrangements in Prostate Cancer: Implications to Genomic Targeting Therapeutics. I. Tereshchenko1,2,3, H. Zhang4, U. Santhanam4, W. Petrosky5, N. Kane-Goldsmith4, J.A. Tischfield2, R.S. DiPaola1. 1) The Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ; 2) Department of Genetics, Rutgers University, Piscataway, NJ; 3) University of Texas MD Anderson Cancer Center, Houston, TX; 2) Harvard Medical School, Boston, MA; 3) Dartmouth College, Lebanon, NH.

Genome-wide association studies have identified several genes that are associated with melanoma risk. However, most of these variants identified by those studies have only weak effects on melanoma risk, as well as on pigmentation or nevi. We hypothesized that the joint analysis of multiple single-nucleotide polymorphisms (SNPs) may detect a larger effect and improve the predictive value of models using standard phenotypic risk factors. In this study, we analyzed 11 SNPs that were associated with melanoma risk in previous studies and were also genotyped in The University of Texas MD Anderson Cancer Center melanoma case-control study and The University of Texas MD Anderson Cancer Center melanoma case-control study and the Health Professionals Follow-Up studies. All 11 SNPs were replicated in the MD Anderson study, but only 5 in the Nurse Health study and 2 in the Health Professionals Follow-Up study were confirmed. Participants who carried 15 or more risk alleles were more than five times as likely to have melanoma as compared to those carrying ≤6 risk alleles in the MD Anderson study. A weighted polygenic risk score (PRS) was constructed using the 11 SNPs to evaluate their joint effect on melanoma risk. In a data set pooled from all 3 studies, per unit increase of PRS led to 1.12 times more likely to have melanoma than those in the lowest PRS tertile (95%; CI, 1.06-1.18; P = 6.63×10-5); Individuals in the highest PRS tertile were 1.69 times more likely to have melanoma than those in the lowest PRS tertile (95%; CI, 1.28-2.25; P = 2.24×10-4), after adjustment for age, sex, and pigmentation (95%; CI, 1.28-2.25; P = 2.24×10-4). These data are extremely important and urgently require further study to understand heterogeneity of genetic changes, because clinical studies have already been launched using genetic changes such as TMPRSS2-ERG fusions found in a single core biopsy to direct a specific therapy.

3414F
Joint effect of multiple common SNPs predicts melanoma susceptibility in place of or beyond traditional risk factors. S. Fang1, J. Han2, M. Zhang1, L. Wang1, Q. Wei1, C. Amos2, J. Lee1. 1) University of Texas MD Anderson Cancer Center, Houston, TX; 2) Harvard Medical School, Boston, MA.

3415T
Trisomy of chromosome 8 in children with haematological malignancies. D. Januszkiewicz-Lewandowska1,2, O. Zajac-Spychala2, E. Malyst2, J. Nowak2. 1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) Department of Pediatric Oncology, Hematology and Transplantation, University of Medical Sciences, Poznan, Poland; 3) Departament of Medical Diagnostic, Poznan, Poland.

Trisomy of chromosome 8 is one of the most common numerical aberrations in adults with AML and MDS, which is present in 5-20% of the cases. Much less frequently is observed in ALL and CML (up to 5% of patients). According to the IPSS cytogenetic classification of MDS trisomy of chromosome 8 is an indelible risk marker. There is no clear data on the prevalence and role of trisomy 8 in pediatric haematological malignancies. In 2009-2012 trisomy of chromosome 8 was found in 5 (4.4%) patients out of 113 children (82 ALL, 26 AML and 5 MDS) treated at the Department of Pediatric Oncology, Hematology and Transplantation in Poznan. In two of them were diagnosed ALL (2/82 - 2.4%), in one AML-M4/M5 (1/26 - 3.8%) and in other to MDS (2/5 - 40%). In three children, trisomy 8 was isolated aberration, in the remaining two, accompanied by monosomy of chromosome 7 (children with MDS) or (9;22), transcript p210/230 BCR/ABL (child with ALL). In 4 patients (one child with ALL was treated only by ALL IC BFM 2002), out of chemotherapy, treatment included HSCT. Only one boy with relapsed leukemia with transformation of primary pre T-ALL to AML-M0 failed treatment. In general observed frequency of trisomy of chromosome 8 in haematological malignancies in children (4.4%) appears to be lower as compared to adults. The decision about the early use of HSCT in the treatment of studied children could be the cause of received good treatment results.

3416F
Radiogenomics: Using Genetics to Identify Cancer Patients at Risk for Development of Adverse Effects Following Radiotherapy. S.L. Krens1,2, H. Oster3,4, B.S. Rosenstein1,5,6. Radiogenomics Consortium. 1) Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Pathology, Albert Einstein College of Medicine, Bronx, NY; 3) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 4) Department of Radiation Oncology, New York University School of Medicine, New York, NY; 5) Departments of Dermatology and Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Radiogenomics (RGx) research aims to identify genetic predictors of response to cancer radiotherapy. RGx focuses on common clinical end-points such as fibrosis or edema associated with radiation treatment and adverse toxicities from radiotherapy for pelvic cancers. It parallels pharmacogenomics with respect to clinical aims and research methods. RGx is a relatively young discipline with many opportunities to develop reproducible methods to determine genetic variation in response to radiotherapy. The field has a strong commonality with the fields of pharmacogenetics and pharmacogenomics which have been enabled by formation of the NCI-supported Radiogenomics Consortium (http://epi.grants.cancer.gov/Consortia/single/rngx.html), which aims to foster collaboration and encourage translation to clinical practice.

3417W
3416T
Posters: Cancer Genetics

Radiogenomics: Using Genetics to Identify Cancer Patients at Risk for Development of Adverse Effects Following Radiotherapy. S.L. Krens1,2, H. Oster3,4, B.S. Rosenstein1,5,6. Radiogenomics Consortium. 1) Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Pathology, Albert Einstein College of Medicine, Bronx, NY; 3) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 4) Department of Radiation Oncology, New York University School of Medicine, New York, NY; 5) Departments of Dermatology and Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Radiogenomics (RGx) research aims to identify genetic predictors of response to cancer radiotherapy. RGx focuses on common clinical end-points such as fibrosis or edema associated with radiation treatment and adverse toxicities from radiotherapy for pelvic cancers. It parallels pharmacogenomics with respect to clinical aims and research methods. RGx is a relatively young discipline with many opportunities to develop reproducible methods to determine genetic variation in response to radiotherapy. The field has a strong commonality with the fields of pharmacogenetics and pharmacogenomics which have been enabled by formation of the NCI-supported Radiogenomics Consortium (http://epi.grants.cancer.gov/Consortia/single/rngx.html), which aims to foster collaboration and encourage translation to clinical practice.

Radiogenomics: Using Genetics to Identify Cancer Patients at Risk for Development of Adverse Effects Following Radiotherapy. S.L. Krens1,2, H. Oster3,4, B.S. Rosenstein1,5,6. Radiogenomics Consortium. 1) Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Pathology, Albert Einstein College of Medicine, Bronx, NY; 3) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 4) Department of Radiation Oncology, New York University School of Medicine, New York, NY; 5) Departments of Dermatology and Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Radiogenomics (RGx) research aims to identify genetic predictors of response to cancer radiotherapy. RGx focuses on common clinical end-points such as fibrosis or edema associated with radiation treatment and adverse toxicities from radiotherapy for pelvic cancers. It parallels pharmacogenomics with respect to clinical aims and research methods. RGx is a relatively young discipline with many opportunities to develop reproducible methods to determine genetic variation in response to radiotherapy. The field has a strong commonality with the fields of pharmacogenetics and pharmacogenomics which have been enabled by formation of the NCI-supported Radiogenomics Consortium (http://epi.grants.cancer.gov/Consortia/single/rngx.html), which aims to foster collaboration and encourage translation to clinical practice.

Radiogenomics (RGx) research aims to identify genetic predictors of response to cancer radiotherapy. RGx focuses on common clinical end-points such as fibrosis or edema associated with radiation treatment and adverse toxicities from radiotherapy for pelvic cancers. It parallels pharmacogenomics with respect to clinical aims and research methods. RGx is a relatively young discipline with many opportunities to develop reproducible methods to determine genetic variation in response to radiotherapy. The field has a strong commonality with the fields of pharmacogenetics and pharmacogenomics which have been enabled by formation of the NCI-supported Radiogenomics Consortium (http://epi.grants.cancer.gov/Consortia/single/rngx.html), which aims to foster collaboration and encourage translation to clinical practice.
3418T

RLIP76 expression levels in CML patients. O. Atikok Clark1, A. Toyul1, O. Salim1, A. Timuragatagolu1, M. Okur3, O.K. Yucel5, N. Sayin Ekin2, M. Ulubas1, K. Eker Gulser1, I. Karadogan1, L. Undar2. 1) Department of Medical Genetics, Akdeniz University Medical Faculty, Antalya, Turkey; 2) Department of Hematology, Akdeniz University Medical Faculty, Antalya, Turkey; 3) Laboratory of Hematology, Akdeniz University Hospital, Antalya, Turkey.

The Ral-interacting protein (RLIP76) is a member of the non-ABC transporter group of multiple drug-resistance proteins, involved in the export of intracellular drugs, GSH conjugates, and lipid peroxidation end products out of the cell. While it is believed that this protein might be responsible for the drug resistance observed in various types of cancers, the expression of RLIP76 has yet to be studied in hematologic malignancies. Furthermore, while it is known that the RLIP76 gene is expressed in the HL60 and K562 leukemia cell lines, the role of the RLIP76 protein in both the pathology of leukemia and chemotherapeutic resistance remains largely unknown. As the prevalence of the drug resistance phenotype, and the resultant grim prognoses, both increases among leukemia patients, the need to better understand the role of the RLIP76 protein is imperative. It was thought that the expression levels of RLIP76 might be important for giving accurate prognoses for CML cases that can’t otherwise be explained by any of the other known drug-resistance genes. Therefore, based on the potential at RLIP76 to be a new and useful prognostic marker for investigating drug-resistance in CML, the expression levels of the RLIP76 gene were measured in CML patients that were receiving imatinib, in this study. Analysis of RLIP76 resistance in CML, the expression levels of the RLIP76 gene were measured other known drug-resistance genes. Therefore, based on the potential at RLIP76 to be a new and useful prognostic marker for investigating drug-resistance in CML, the expression levels of the RLIP76 gene were measured in CML patients that were receiving imatinib, in this study. Analysis of RLIP76 expression levels in CML patients. The presence of circulating microRNAs in bone marrow of ALL patients shows their potential as tumor microenvironment regulators and biomarkers.

3419F


Circulating microRNAs have been found in all body fluids and could be associated with modulation of tumor microenvironment and progression of diseases. In the presence of cancer cells, active components present in normal microenvironment, become deregulated, promoting cell survival, disease progression and drug resistance. Several microRNAs act as participants in the process of development of leukemia either increasing or decreasing the expression levels. Here, we show data where circulating microRNAs were found in bone marrow serum from acute lymphoblastic leukemia (ALL) patients. We investigated, by smallRNAseq, the presence of microRNAs in tumor microenvironment from four patients, collected at diagnosis (phase one), during chemotherapy treatment (phase two and three). After statistical analysis, only microRNAs with expression difference between treatment phases with a p value under 0.05 (p<0.05) were considered. In totally, 946 known microRNAs were identified, 83 presented significant difference expression between phase one and two, 57 between phase one and three and 56 comparing phase two to three. miR-146a, identified in disease progression, with expression inversely proportional to patient survival, appeared with reads count decaying through chemotherapy phases, suggesting a reduction in the expression within cells and subsequent reduction in the microenvironment, miR-181, described in studies as participating in leukemia development also presented significativity alteration during chemotherapy treatment. The presence of circulating microRNAs in bone marrow of ALL patients shows their potential as tumor microenvironment regulators and biomarkers.

3420W

microRNA expression profiling in metastatic cutaneous squamous cell carcinoma. A.E. Toland1,2,3, L.E. Skeele3, S.B. Peters3, T. Teknos3, T. Olenczi3,2, O.C. Aliain1,2,1) Human Cancer Genetics and Comprehensive Cancer Center, Ohio State University, Columbus, OH; 2) MVIMG, Ohio State University, Columbus, OH; 3) OSU Wexner Medical Center, Columbus, OH.

Cutaneous squamous cell carcinoma (cSCC) is the second most common form of cancer with approximately 700,000 cSCCs diagnosed in the United States each year. The majority of cSCCs are readily treated surgically; however, 2-6% of all cSCCs metastasize leading to approximately 2500 deaths annually. Metastatic cSCC is associated with a survival rate of 56% with distant metastases and a worse outcome in recurrent cSCCs and immunocompromised individuals. No published studies have been performed to date to assess microRNA (miRNA) profiles of metastatic cSCC samples. In order to develop specifically targeted therapies for metastatic cSCC and to determine which factors are predictive of metastatic cSCC from early stage lesions, it is crucial to understand the molecular events which lead to metastasis of cSCC. We hypothesized that aberrant expression of miRNAs enhances the metastatic potential of cSCCs. To test this hypothesis, we performed miRNA expression analysis using the nCounter miRNA panel of approximately 800 miRNAs in 48 samples including 10 trips of normal, primary tumor and metastatic samples and 9 paired normal/non-metastatic primary cSCCs. We identified ~225 miRNAs expressed in the skin and performed analysis on the top 100 expressed miRNAs. Multiple miRNAs showed significant differences between primary cSCCs that did not metastasize compared to metastatic cSCCs or their corresponding primary tumor. These include miR-let-7f, miR-4286, miR-15, miR-16, miR-29a and miR-29b; all of which have been previously associated with metastatic cancers, suggesting the validity of these results. Interestingly, miR-4286 is up-regulated in metastatic melanoma and shows a 6-7-fold up-regulation in our metastatic cSCC samples. miR-21 has been shown to be associated with metastasis in multiple tumors including esophageal SCC. Differentially expressed miRNAs are currently being evaluated in a second cohort of cSCCs and metastatic samples. In summary, several miRNAs show differential expression between non-metastatic and metastatic cSCCs; these may be useful as biomarkers for cSCC metastasis or as targets for therapy.

3421T

Prognostic significance of K-ras codon 12 mutation with resected Gall Bladder Cancer. H.R. Kazmi1, A. Chandra1, D. Parmar2, N. M1. 1) Surgical Gastroenterology, King George’s Medical University, Lucknow, Uttar Pradesh, India; 2) Developmental Toxicology Division, Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India.

Introduction: Prognosis of Gall bladder cancer (GBC) is dismal due to aggressive behavior and lack of effective treatment. High incidence and mortality rates have been observed in Northern India. In present study, we had analyzed K-ras mutation in normal (undiseased) and GBC tissue samples and investigated its prognostic significance. Materials and Methods: Histopathologically proven GBC (n=39, stage II: n=22 and stage III: n=17) and normal gall bladder (n=24) tissue samples were studied. DNA was isolated and mutation was detected by Restriction Fragment Length Polymorphism analysis. Statistical analysis was carried out using chi-square test with Yates correction and Fisher exact test. The effect of K-ras mutation on patient survival was estimated using the Kaplan-Meier method and the difference between curves was analyzed by log-rank test. Results: The mean age (years + S.D.) of normal and GBC patients were 41.41±13.42 and 43.87±12.39 (range 22-67) respectively. K-ras codon 12 mutation was observed in 1/24 (4.17%) and 16/39 (41.03%) of normal and GBC tissue samples respectively and with statistically significant difference (p=0.001). The overall survival time was significantly shorter (p=0.003) in patients having K-ras mutation. The median survival for GBC patients was shorter with K-ras mutation as compared to patients without K-ras mutation (12.5 vs. 17 months). Statistically significant difference in overall survival was also observed for stage II (p=0.012, median survival -15 months) and Stage III (p=0.009, median survival -8 months) GBC patients with K-ras mutation. Overall K-ras mutation significantly affects prognosis for GBC (Hazard ratio [HR] = 5.54; 95% Confidence interval [CI] = 1.54-18.14). For stage II GBC, K-ras positive patients carry 4.3 times higher risk of failure as compared to K-ras negative cases (95% CI = 1.37-13.55). Similarly, the HR for stage III disease was 7.42 (95% CI = 1.66-33.11). Conclusion: High frequency of K-ras codon 12 mutation in Northern Indian population. Its presence indicates poorer outcomes in patients undergoing surgery for this disease.
Posters: Cancer Genetics

3422F

"Genetic polymorphisms in the apoptotic-associated genes Fas (-670
ag) and Fas L (-844 tc) & acute myeloid leukemia risk". Vuree. Sugunakar 1, C. Anuradha 1, Dunna. Nageswara rao 4, G. Manjula 1, EM. Prajitha 1,
D. Raghnadharao 3, Sinha. Sudha 2, Atili. Venkat Satya Suresh 2, Satti. Vishnupriya 1. 1) Genetics and Biotechnology, Osmania University, Hyderabad,
Andhra Pradesh, India; 2) MNJ Institute of Oncology, Red Hills, Hyderabad;
3) NIMS, Panjagutta, Hyderabad; 4) Sastra University, Thanjavur, Tamilnadu.
Acute Myeloid Leukemia (AML) is a cancer of myeloid lineage of blood
cells, characterized by the rapid growth of abnormal white blood cells that
accumulate in the bone marrow and interfere with the production of normal
blood cells. FAS (TNFSf6/CD95/APO-1) belongs to the family of tumor
necrosis factor receptors, and binding to the receptor by the FAS ligand
(CD95L) triggers receptor trimerization and subsequent assembly of the
death-inducing signaling complex. This gene is silenced in many tumor
types, resulting in an inability to respond to proapoptotic signals. The FAS
promoter is polymorphic, including an A to G substitution at -670bp, and T
to C substitution at -844bp, which occur within SP1 and signal transducers
and activators of transcription 1 transcription factor binding sites, respectively. The present study on the association of Fas Receptor (Fas R) and
Fas Ligand (Fas L) promoter polymorphisms with AML were analyzed by
PCR-RFLP method. GG genotype of Fas -670 AG showed a significant
association with respect to the AML as well as with epidemiological variables
like Sex of the proband, Area of living (Urban), WBC (30,000), no significant
trend was observed with platelet count and Complete Remission. Whereas
in FasL-844TC, C allele of FasL-844 was elevated in cases when compared
with controls. Further, MDR analysis was performed between the two promoter SNPs which showed a significant trend.

3423W

TRAPing Telomerase Activity Using Droplet Digital PCR (ddPCR). D.N.
Shelton 1, J. Lin 2, J.F. Regan 1, G. Karlin-Neumann 1, E.H. Blackburn 2. 1)
Bio-Rad Laboratories, Digital Biology Center, Pleasanton, CA; 2) UCSF
School of Medicine, Dept. of Biochemistry and BioPhysics, San Francisco, CA.
The aim of this work was to develop a more sensitive and high throughput
assay for measuring telomerase activity. Telomeres are the protective structures at the ends of chromosomes consisting of 6 bp repeat sequences. In
young cells, these regions can be as long as 15kb and act as caps which
protect the DNA ends. These ends naturally degrade with each passing
cell division, usually losing 25-200 base pairs per division. Once they are
shortened below a critical length (estimated to be 200-300 bp) the cells
arrest and become senescent, or ‘old’. Telomeres can be thought of as a
cellular or mitotic clock. Once the clock has wound down, the cells either
die or pass through crisis and become immortal. One of the mechanisms
of immortality is the activation of the enzyme, telomerase. Telomerase is
the endogenous reverse transcriptase responsible for adding repeats to
telomeres, rewinding the clock and enabling a cell to continuously divide.
Abundant telomerase activity is found in the majority of cancers, fetal and
adult stem cells, and germ cells. It is also present at much lower levels in
non-pluripotent cells, such as immune cells, but these levels are currently
difficult to measure. The telomerase repeat amplification protocol (TRAP)
measures the presence of active telomerase by measuring the activity of
the enzyme on a starting template, which is then amplified by PCR. For
samples with abundant telomerase activity, SYBRGreen® qPCR assays
provide high throughput. However, the current most sensitive method of
detection still uses radioactivity and laborious PAGE sequencing gels followed by densitometry to quantify telomerase. In this study, the PAGE
detection method is replaced by single molecule counting of telomeraseextended templates using droplet digital PCR technology.

3424T

633

Splicing of HRAS exon 2 is vulnerable - The splicing efficiency of
activating mutations in codons 12 and 13 determines Costello syndrome phenotype. B.S. Andresen 1, A.M. Hartung 1, J. Swensen 2, I.E. Uriz 1,
M. Lapin 1, J.C. Carey 3, A. Calhoun 4, P. Yu 2, C.P. Vaughn 2, S.F. Dobrowolski 5, M.R. Larsen 1, H. Hanson 3, D.A. Stevenson 3. 1) Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense,
Denmark; 2) University of Utah School of Medicine, ARUP Laboratories,
Salt Lake City, Utah; 3) Division of Medical Genetics, University of Utah
School of Medicine, Salt Lake City, Utah; 4) Division of Pediatric Genetics
and Metabolism, University of Minnesota, Minneapolis, Minnesota; 5)
Department of Pathology, Children's Hospital of Pittsburgh, Pittsburgh,
Pennsylvania.
Costello syndrome (CS) is most frequently caused by a c.34G>A (p.G12S)
activating mutation in HRAS with modest transforming activity. G12V mutations have the highest transforming activity, the greatest frequency in cancers, but are very rare in CS. So far, all CS patients with p.G12V mutations
(c.35G>T; c.35_36delinsTT; c.35_36delinsTA), have had a severe, early
lethal, phenotype. Sequence analysis of a 12-year-old boy with an attenuated
CS phenotype revealed, to our surprise, a new germline p.G12V mutation,
c.35_36delinsTG, without evidence of mosaicism. In silico analysis shows
that exon 2 has a weak 3’ splice site and that c.35_36delinsTG simultaneously abolishes exonic splicing enhancer (ESE) motifs and creates exonic
splicing silencer (ESS) motifs indicating that it may disrupt splicing. Analysis
of patient HRAS cDNA showed that c.35_36delinsTG results in exon 2
skipping and consequently little mutant protein, explaining the attenuated
phenotype. Transfection of several different cell lines with HRAS mini genes
with the four known p.G12V mutations and the common c.34G>A mutation
showed that only c.35_36delinsTG results in exon 2 skipping. Deletions of
nucleotides c.32-37 cause exon 2 skipping, indicating presence of an ESE
in this region. Optimization of the weak 3’ splice site corrected splicing from
the mutants confirming vulnerability of exon 2. Testing of wild type and
several mutant HRAS exon 2 sequences in different splicing reporter minigenes showed that ESE strength is increased by the other p.G12V mutations
and by other mutations with strong transforming potential, whereas c.35_
36delinsTG inactivated splicing. RNA affinity purification, ITRAQ labeling
followed by MS/MS, showed that c.35_36delinsTG increases binding of
hnRNPF/H splicing inhibitory proteins, whereas c.35G>T increased binding
of several splicing stimulatory SR proteins consistent with the observed
effect on splicing. Replacement in the HRAS minigene of the wild type
sequence with hnRNPF/H binding ESS motifs confirmed that binding of
hnRNPF/H results in exon 2 skipping. Our study illustrates that phenotype
of Costello syndrome and likely somatic cancers are not only determined
by the transforming potential of the mutant HRAS protein, but is also determined by the efficiency of exon 2 inclusion. This has important implications
for our understanding of the correlation between genotype and phenotype in
diseases caused by HRAS mutations and for development of new therapies.


3426W
DICER1 Mutations in Pituitary Blastaoma: new gene, new disease. L. de Kock1-3, N. Sabbaghian4, F. Plourde4, A. Srivastava5, D. Bouron-Dal Soglio2, N. Hammoud6, J.H. Choi7, S. Park2, C.L. Deay7, J.I. Dhop3, A. Estenssore4, T. Jacques10,11, A. Perry7, H. Leichter13, P. Maeder14, M.A. Brundler15,16, J. Neal17, M. Zacharin18, M. Korbonits19,20, T. Cole21, S. Albrecht22, E. Horvath23, K. Kovacs23-25, J.R. Priest26, W.D. Foukles2, J. Mc Gill University, Montreal, QC, Canada; 5) Department of Pathology, Yeungnam University College of Medicine, Daegu City, South Korea; 6) Seoul National University College of Medicine, Seoul, Republic of Korea; 7) Department of Endocrinology, CHU Ste-Justine, Montreal, QC, Canada; 8) Department of Pathology and Laboratory Medicine, Children's Hospital Colorado, USA; 9) Department of Pediatrics, Monroe Carell Jr. Children's Hospital, Nashville, TN, USA; 10) National Development Unit, UCL Institute of Child Health, London, UK; 11) Department of Histopathology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK; 12) Departments of Pathology and Neurological Surgery, UCSF Medical Centre, San Francisco, CA, USA; 13) Department of Pediatrics, Olghospital, Dresden, Germany; 14) Department of Neuroradiology, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 15) University of Calgary, Alberta, Canada; 16) Alberta Children's Hospital, Alberta, Canada; 17) Department of Histopathology, UHW, Cardiff, UK; 18) Department of Endocrinology and Diabetes, Royal Children's Hospital, Flemington Road, Parkville, Australia; 19) Department of Endocrinology, Barts and the London School of Medicine, London, UK; 20) Queen's Mar United Hospitals, London, UK; 21) Birmingham Women's NHS Foundation Trust, Edgbaston, Birmingham, West Midlands, UK; 22) Department of Pathology, Montreal Children's Hospital, McGill University Health Centre, Montreal, QC, Canada; 23) Department of Laboratory Medicine and Pathobiology, St. Michael's Hospital, University of Toronto, Toronto, Canada; 24) Minneapolis, Minnesota, USA.

BACKGROUND: DICER1, a non-coding small RNA processing enzyme, cleaves micro RNA (miRNA) precursors into mature miRNAs, which are known to regulate gene expression. Mutations in DICER1 have recently been found to predispose to a rare cancer syndrome consisting of primarily childhood tumours. The main manifestations of the syndrome include pleuropulmonary blastoma (PPB), cystic nephroma (CN), Sertoli-Leydig cell tumours (SLCT), multinodular goiter and other rare childhood sarcomas and dysplasias. We and others have shown that germ-line DICER1 mutations are often accompanied by specific somatic mutations in the DICER1 RNase III catalytic domain in several tumour types, and we hypothesized that both germ-line and somatic mutations might also occur in children with pituitary blastoma (PitB). PitB, identified as a distinct entity in 2008, is a very rare, potentially lethal childhood brain tumour. Since the discovery by our team of an inherited mutation in DICER1 in a child with PitB in 2011, we have identified a further 12 pathologically-confirmed PitB cases. Histological studies suggested perturbation of pituitary stem cells could underlie the pathogenesis of PitBs. OBJECTIVES: We aim to determine the contribution of germ-line and somatic DICER1 mutations to PitB and to analyse the effect of the mutations on the mi/mRNA landscape. We hypothesize that targets involved in the regulation of pituitary stem cells will be dysregulated in these tumours.

RESULTS: To date, 6 of 6 PitB cases analysed genomically harbour a germ-line DICER1 mutation, suggesting that mutation of DICER1 is a key predisposing genetic event. Furthermore, we identified likely contributory “second hits” in DICER1 in 6 of the 7 PitB cases tested thus far for somatic mutations. Five of these 6 mutations occur within the RNase IIIb catalytic domain, a domain essential to the generation of miRNAs from the 5′ arm (5p) of precursor miRNAs. Work is underway to complete the analysis of the remaining samples and to determine the effect of these DICER1 mutations on the generation of miRNAs. CONCLUSION: Germ-line DICER1 mutations are a major contributor to PitB. Second somatic ‘hits’ in DICER1, within the RNase IIIb domain, also appear to be critical steps in their pathogenesis. Further studies will focus on miRNA, mRNA and gDNA profiling of these rare tumours.

3425F
Familial intracranial meningioma without NF2 and SUFU mutations: searching for additional predisposing genes. M. Aavikko1, H. Ristola- lainen1, E. Kaasinen1, P.A. Koivisto2, M. Pöyhönen3, K. Claess4, T. Van Maaiken5, M. Artama6, E. Pukkala7, L.A. Aaltonen8, P. Vahteristo9. 1) Department of Medical Genetics, Genome-Scale Biology Research Program, University of Helsinki, Helsinki, Finland; 2) Seinäjoki Central Hospital, Department of Neurology, Seinäjoki, Finland; 3) Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 4) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 5) The Finnish Cancer Registry, Helsinki, Finland.

Meningiomas are the most common tumors of the central nervous system (CNS) accounting for one third of all primary CNS tumors in adults. Majority of meningiomas are slowly growing benign tumors arising from the leptomeninges covering CNS. Most commonly they occur sporadically, but they are also commonly seen in individuals with Neurofibromatosis type 2 tumor suppressor syndrome.

We recently reported a candidate predisposing gene defect in the Suppres sor of Fused homolog (SUFU) in a family of five siblings with intracranial meningiomas, four of whom had multiple lesions. All seven studied meningiomas from the family displayed loss of the wild type allele following the two-hit model for tumor suppressor genes. In addition, we showed that the mutant SUFU had lost its ability to bind GLI1 transcription factor causing dysregulated Hedgehog signaling. Mutations in SUFU have previously been identified in familial and sporadic medulloblastomas. Our results suggested that in addition to medulloblastomas SUFU defects may also predispose to a subset of meningiomas.

Here we describe a Finnish NF2 and SUFU mutation negative family of five affected individuals with intracranial meningiomas. The family display autosomal dominant inheritance pattern with affected individuals in three generations. We have performed genome-wide SNP genotyping and linkage analysis in combination with exome and genome sequencing. We have identified candidate chromosomal regions and gene variants in this family and we are studying the pathogenicity of these variants in the meningioma tumors from the family and in additional familial and sporadic intracranial meningioma cases, these include a Belgian family with four affected individu als in four generations and Finnish familial and sporadic meningioma patients with intracranial meningiomas. The family display a subset of meningiomas.
3427T
Targeted sequencing of MLH1, MLH2, and MSH6 in defective mismatch repair colorectal cancer cases with no identified mutation. M. De Rycke1, S. Gunawardena2, S. McDonnell3, S. Siddha4, S. Riska1, Z. Fogarty1, B. Eckloff1, D. Schaid1, E. Goode1, N. Linder4, S. Thibodeau1, Colon Cancer Family Registry. 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Advanced Genomics, Foundation Medicine, Mayo Clinic, Rochester, MN; 4) Health Sciences Research, Mayo Clinic, Scottsdale, AZ.

Approximately 25% of colorectal cancers (CRC) are clustered in families expected to harbor predisposing genetic mutations. Of all CRC cases, ~3-5% are caused by Lynch Syndrome (LS) in the United States and 5-10% in Europe. MLH1, MLH2, MSH2, MSH6, and PMS2. Currently, when LS is suspected, germline testing of the proband can be done to identify the underlying mutation and facilitate testing relatives who may be at increased risk. Typically, the entire coding region and exon/intron boundaries of the gene are targeted for sequencing, and identified variants found are classified as deleterious, variants of uncertain significance (VUS), or common polymorphisms. A significant fraction of cases show loss of protein expression by immunohistochemistry but no detected variant or deletion in the affected gene, presenting a problem for families with suspected LS: no mechanisms exist to determine which family members share the variant and are at increased risk for developing CRC. In this study, we completed targeted sequencing of MLH1, MLH2, and MSH2 in 130 individuals with defective MMR CRC in which previous sequencing studies failed to detect any disruptive germline MMR variants. We targeted exonic, intronic, and nearby intergenic regions of MLH1, MSH2, MSH6, as well as the 3' end of EPACAM. Agilent's SureSelect Custom Capture system was used to target regions of at least 10-20 variants with at least 10x coverage by sequencing on a HiSeq 2000. Samples underwent extensive quality control, which resulted in the exclusion of 10 samples due to low coverage (n=7) or high read duplication (n=3). Mean coverage for the remaining 120 samples was 633x (range 164 - 3,114x; median 367x). Transition-Transversion ratios for known and novel SNPs were 2.34 and 2.29, respectively. Samples had an average of 3,124 SNPs (range 2,639 - 4,168) and 386 (range 249 - 464) INDELs in the capture region. Previously completed genotyping results were compared with sequencing results for 118 samples, resulting in a 96% concordance. Preliminary analysis of the 85 cases with loss of MLH1 protein expression identified the -93G>A transition in 45 cases, with six individuals being homozygous recessive. This variant has been reported to reduce expression of MLH1 and has been considered a low penetrance mutation contributing to CRC development. Further results from the analysis will be presented.

3429F
Anotation of Rare Variants from Exome Sequencing in Families with Lymphoid Malignancies. L.R. Goldin1, M.L. McMaster2, M. Rotunno2, J. He2, L. Burdette3, A. Hutchinson3, J. Boland3, M. Yeager3, M.A. Tucker4, S.J. Chanock5, N.E. Caporaso3. 1) Genetic Epidemiology Branch, DCGE/NCI, Bethesda, MD; 2) Cancer Genetics Research Laboratory, DCGE/NCI, Bethesda, MD; 3) Human Genetics Program, DCGE/NCI, Bethesda, MD.

Several variants with relatively small effects have been identified from GWAS of lymphoid malignancies but single high penetrance genes have not been identified. We have conducted exome sequencing in 176 individuals from 49 families at high risk for lymphoid malignancies including chronic lymphocytic leukemia, Hodgkin lymphoma, and non-Hodgkin lymphomas. We used Nimblegen v.2.0 and v3.0 for exome capture followed by sequencing on the Illumina HiSeq 2000. We required that 80% of coding sequences had at least 15x coverage. Novoalign v.2.07.14 was used for alignment and GATK was used for local re-alignment and variant calling. We sequenced 3 or more patients or obligate carriers from the families. A large number (~1000) of rare (<1% frequency in European populations) rare variants were shared among patients/carriers in each family. In order to eliminate other possible sequencing artifacts, variants found in more than 1% of samples sequenced in our laboratory from other studies were excluded, leaving ~20-100 variants per family. We prioritized variants for further validation and follow-up by looking for genes with the same shared variant in more than one family that were predicted to be damaging by at least one program (or were in a regulatory region), were highly conserved, and had a literature link to cancer or a cancer-related pathway. The program, Ingenuity Variant Analysis, was used to facilitate filtering and annotation. Variants in 7 genes (BORA, FAF1, ITGB2, PRDM2, TXNDC17, ZNF189, and ELMO3) met these conditions. We will conduct targeted sequencing of these genes in 366 patients from an additional 170 multiplex families. There were an additional 20 families with different interventional and 55 samples from 10 additional families with greater than one family and selected genes from this group will also be sequenced in the new sample. Results of our validation/follow-up studies will be presented.

3429W
Constitutional mismatch repair deficiency syndrome caused my MMR gene founder mutations with a high prevalence in Israel. I. Kedar-Barnes1, E. Karp1, E. Halfi2, M. Halpern3, I. Lerner3, T. Peretz3, P. Kariv2, E. Halfi1, N. Magal1, K. Wimmer2, Y. Goldberg4, D. Berkovich5, H.N. Baris1, 1) The Raphael Recanati Genetics Inst, Rabin Med Ctr, Petakh Tikva, Israel; 2) Gastroenterology Division, Rabin Medical Center, Bellinison Hospital, Petakh Tikva, Israel; 3) Pediatric Oncology, Schneider Children’s Medical Center, Petach Tikva, Israel; 4) Department of Pathology, Rabin Medical Center, HaSharon Hospital, Petakh Tikva, Israel; 5) Department of Human Genetics, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; 6) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 7) Department of Gastroenterology & Liver Disease, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; 8) Gastroenterology Department, Rambam Health Care Campus, Haifa, Israel; 9) Division of Human Genetics, Medical University Innsbruck, Austria; 10) GGA-Genetik Analyse, Kazerin, Israel.

Heterozygous germline mutations in one of the 4 mismatch repair (MMR) genes, MLH1, MSH2, MSH6 and PMS2 cause Lynch syndrome (LS), an autosomal dominant cancer predisposition syndrome conferring a high risk of colorectal, endometrial and other cancers in adulthood. Offspring of couples, both having LS, have a high risk to inherit biallelic MMR gene mutations. The cause constitutional MMR deficiency (CMMRD), a severe recessively inherited childhood cancer syndrome with a very broad tumor spectrum including mostly hematological malignancies, brain tumors and childhood colon cancer. Most of CMMRD children also present with early life events such as infections mimicking Neurofibromatosis type 1. We describe our experience in Israel with 5 CMMRD families. The clinical presentation of CMMRD includes: brain tumors at age 3-10 years, two or more patients with lymphoma at age 10-20 years, and one patient with lymphoma at age 12. In two non-consanguineous Ashkenazi families, the common founder Ashkenazi mutation were detected: one family was homozygous for c.1906G>C in MSH2 and the 2nd family heterozygous for c.3984_3986dup3CTCA and c.3859_3862delCAAG in MSH6. In the 3 other consanguineous families, different homozygous mutations were identified: c.2192T>G (p.L731X) in PMS2, a recurrent mutation among Iranian Jews, PMS2 c.686_687delCT and MSH6 3603_3606delCT mutations were found in large pedigrees of Bedouin and Arab ancestry, respectively. Given the prevalence of these mutations among Israeli population, we want to raise the awareness of CMMRD syndrome and open a discussion regarding screening for MMR founder mutations among spouses of LS patients for the purpose of prenatal diagnosis.

3430T
Magnetic Resonance Imaging screening In Li Fraumeni Syndrome: An exploratory whole body MRI study (the SIGNIFY study). E. Killick1,2, E. Bancroft1, N. Taylor2, D.G. Evans3, M. Leach1,3, R. Eeles4,5,6,7, The SIGNIFY collaborators and steering committee. 1) Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) Cancer Care, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom; 3) Royal Marsden NHS Foundation Trust, Sutton, Surrey, United Kingdom; 2) Cancer Care, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom; 3) Royal Marsden NHS Foundation Trust, Sutton, Surrey, United Kingdom; 2) Cancer Care, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom; 3) Royal Marsden NHS Foundation Trust, Sutton, Surrey, United Kingdom; 2) Cancer Care, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom; 3) Royal Marsden NHS Foundation Trust, Sutton, Surrey, United Kingdom; 2) Cancer Care, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom; 3) Royal Marsden NHS Foundation Trust, Sutton, Surrey, United Kingdom; 2) Cancer Care, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom; 3) Royal Marsden NHS Foundation Trust, Sutton, Surrey, United Kingdom.

Background Li Fraumeni Syndrome predisposes individuals to a range of different malignancies with a lifetime cancer risk of up to 90% in women and 70% in men. Current national screening recommendations are for dual modality breast screening with mammography and MRI in women. Some centres employ family specific screening tailored to malignancies found at increased frequency within certain families with most centres having an ‘open-door’ policy. Recent evidence suggests there may be a survival benefit for more intensive screening, including whole body MRI; there are no published data on the psychological impact of such screening programmes. The primary study end-point is to assess the incidence of malignancies diagnosed in asymptomatic TP53 mutation carriers using whole body MRI, against general population controls. The secondary end-points are to assess the incidence of non-malignant relevant disease, to assess the incidence of irrelevant findings and the investigations required to determine relevance of MRI findings and finally to assess the psychological impact of whole body MRI screening in TP53 mutation carriers. Methodology We will recruit 44 TP53 mutation carriers and 44 population controls aged 18 - 60 years, who will undergo conventional and diffusion weighted whole body MRI. The MRI will be read by two independent radiologists in consensus and incidental findings will be relayed to the individual and any necessary further investigations arranged. An inter-rater virtual MDT will be held to review all the reports and to confirm the incidental findings. A series of questionnaires will be used from time of recruitment up to a year after the MRI scan to assess the psychological impact of screening. Results The study recently opened to recruitment; thus far 4 individual participants have been enrolled. Results from these and further recruits will be shown.
Characterization of RB1 mutations and incidence of undetected defects in retinoblastoma. P.S. Lai1, A.P. Alcasabas1, G. Sundar1, B.L. Quach1, 1Dept Ophthalmic, National University Hospital, Singapore; 2) Dept Ophthalmology, National University Hospital, Singapore; 3) Singapore National Eye Centre, Singapore.

Retinoblastoma (RB) is a retinal tumor associated with biallelic loss of RB1 gene. Most cases are diagnosed by five years of age and occur in both hereditary and sporadic forms. Bilateral RB cases are usually associated with constitutional or germline mutations while approximately 15% of sporadic unilateral cases may carry heritable constitutional mutations. Recently, there has been speculation that some rare RB tumors may arise by mechanisms other than RB1 mutations. In this study, we characterized RB1 mutations in 40 retinoblastoma cases in order to determine the incidence of non-detectable RB1 mutations. To detect large deletions, segregation analysis using polymorphism markers, quantitative multiplex PCR or MLPA assays were carried out while small point mutations were detected by sequencing analysis. RB1 promoter methylation was analysed by methylation-specific PCR. RESULTS: Large deletions were present in 27 out of 40 (67.5%) tumors. Germline mutations were detected in 11 out of 13 bilateral cases (84.6%). Among the 27 unilateral RB cases, germline mutations were detected in 4 cases (14.8%). The mutation spectrum shows RB1 mutations (7 nonsense and 2 frameshift mutations) in 16 cases as distributed among exons 12-22 which carries for large RB protein-poor domain known to be critical for transcriptional repression. The most frequent mutations were nonsense (24/35; 68.6%) followed by frameshifts (7/35; 20%) with splice site and missense mutations contributing towards the remaining genetic defects. Most of the mutations identified in this study correspond to known RB gene defects. No de novo mutation were found in four CGA codons located in exons 1, 18 and 27. These findings correspond with previously published data analysed by meta-analysis. RB1 promoter hypermethylation was rare and observed only in 1 tumor. CONCLUSIONS: Mutation in only one RB1 allele was detected in 5 cases while no RB1 mutations on both alleles were found in 2 unilateral cases. This could be due to (1) undetected translocations, deep intronic mutations, or alterations in unknown RB1 regulatory regions, (2) presence of low level RB1 mosaic mutations not detectable by constitutional DNA, several shared CNVs were identified in tumors. The majority of the RB1 mutations identified in this study correspond to previously published data. No mutations were found in two unilateral cases. The data support genetic heterogeneity of RB1 mutation in retinoblastoma.

Frequencies of BRCA1, BRCA2, PALB2, and CDKN2A germline mutations in familial pancreatic cancer (FPC): A PAGENE study. D.B. Tailandier1, J. Robyn1, N. Rigault2, A.G. Schwandner3, M.G. Goggins4, R.H. Ruban5, M.L. Cote6, K. Moyes7, R.J. Wenskrip8, A.R. Hartmann9, D. Seminara10, A.P. Klein11, G.M. Petersen12, 1) Mayo Clinic, Rochester, MN; 2) Toronto General Hospital, Toronto, Ontario, Canada; 3) Dana-Farber Cancer Institute, Boston, MA; 4) Karmanos Cancer Institute, Detroit, MI; 5) Johns Hopkins University, Baltimore, MD; 6) Myriad Genetic Laboratories, Inc, Salt Lake City, UT; 7) National Cancer Institute, Bethesda, MD.

Background: FPC (defined as at least two affected first degree relatives (FDR) in a kindred) accounts for 5-10% of pancreatic cancers and is thought to be genetically heterogeneous. Genetic testing is available for germline mutations in several genes known to predispose to FPC. However, our knowledge is limited because published studies to date have been small and have been non-systematic.

Methods: The Pancreatic Cancer Genetic Epidemiology Consortium (PAGENE) has assembled a registry of 1747 kindreds containing two or more relatives affected with pancreatic cancer. Five PAGENE sites tested germline DNA samples from 234 undiagnosed cases affected with pancreatic adenocarcinoma (80 met criteria for FPC) for four genes: BRCA1 and BRCA2 (including analysis of deletions and rearrangements), PALB2, and CDKN2A.

Conclusions: This is the first comparative study of four susceptibility genes in a large sample of FPC cases. The data support genetic heterogeneity of FPC. Genetic testing for cases meeting FPC criteria may be warranted, but multiple genes must be evaluated. The genetic basis for the majority of FPC remains to be determined.

Exome Sequencing of Family with Carcinoid Cancers. D.W. Neklason1, A. Snow2, C. Teetlink2, 1) Huntsman Cancer Inst, Univ Utah, Salt Lake City, UT, 84112; 2) Department of Medicine, University of Utah, Salt Lake City, UT, 84112; 3) Division of Genetic Epidemiology, University of Utah, Salt Lake City, UT 84112.

We describe here genetic study of a family with carcinoid tumors of the gastrointestinal tract. Two siblings and a first cousin were diagnosed with carcinoid tumors of the small intestine and appendix, respectively. Carcinoid tumors are neuroendocrine tumors and most neuroendocrine tumors are not familial. Incidence of neuroendocrine tumors is 2.47 per 100,000 with the rate steadily increasing over the past 20 years. The majority of neuroendocrine tumors occur in the GI tract (67%), primarily in the small intestine as represented in this family. Carcinoid tumors derive from enterochromaffin cells of the neuroendocrine system in the gut. These cells contain a large amount of the body’s store of serotonin and in response to stimuli in the lumen (chemical, mechanical, pathological), the serotonin release signals nausea to the brain. The two cases who were tested were not previously known to carry germline mutations in these genes. We also studied mutation frequencies between FPC versus familial non-FPC cases (at least two affected blood relatives, but not FDR).

Results: Frequencies for deleterious/suspected deleterious (DS) mutations varied among cancer susceptibilities (CSC) and were: BRCA1 2/20 (10%), BRCA2 3/27 (11%), PALB2 10/77 (1.3%), no VUS; CDKN2A 4/80 (5.2%), 3 VUS. Among Familial non-FPC cases, frequencies were: BRCA1 0/123, no VUS; BRCA2 1/123 (0.8%), 1 VUS; PALB2 1/147 (0.7%), 4 VUS, CDKN2A 0/153, 3 VUS. FPC cases carried a total of 7, FPC and familial non-FPC cases carried 2 and 1, respectively. We excluded known mutations in our registry from this study, these frequencies may be a lower estimate. No case carried more than one mutation.

Conclusions: This is the first comparative study of four susceptibility genes in a large sample of FPC cases. The data support genetic heterogeneity of FPC. Genetic testing for cases meeting FPC criteria may be warranted, but multiple genes must be evaluated. The genetic basis for the majority of FPC remains to be determined.

Common Somatic Variations Identified in Maffucci Syndrome. M. Amy3,1, A. Palma4, A.P. Klein5, I.Kaitila3, P.L. Docquier6, 1) Department of Dermatology, Hôpital Lariboisière, Paris, France; 2) Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 3) Division of Orthopedic Surgery, Cliniques Universitaires St-Luc, Université catholique de Louvain, Bruxelles, Belgium; 4) Department of Plastic Surgery, Children’s Hospital and Harvard Medical School, Boston, MA, USA; 5) Center for Vascular Anomalies, Division of Plastic Surgery, Cliniques Universitaires St-Luc, Uni- versité catholique de Louvain, Brussels, Belgium; 6) Wallon Excellence in LifeSciences and Biotechnology (WELBIO), Université catholique de Louvain, Brussels, B-1200, Belgium.

Maffucci syndrome (MS) is a rare congenital disorder characterized by multiple cutaneous and subcutaneous enchondromas accompanied by cutaneous spindle-cell hemangiomas. These patients have a high incidence of malignant transformation. No familial case is known and the etiopathogenic cause remains unknown. In enchondromatosis (Ollier’s disease, SD), patients are comprised of enchondromas only, four mutations in the PTH1R gene have been identified in four patients: three were somatic and one was germline. No PTHR1 mutation has been detected in MS. On the other hand, in 77% of patients with MS and 81% of patients with OD, somatic IDH1 and more rarely IDH2 mutations have been observed. These changes are shared with other tumours, including glioblastomas, leukemias and thyroid cancers, and seem to be markers of cellular transformation. To search for underlying somatic genomic causes, we screened MS tissues using Affymetrix SNP-chips. We looked for common copy number variations (CNV), loss of heterozygosity (LOH) and familial rearrangements (germline and somatic). We analyzed two families, one combined in constitutional DNA versus the corresponding blood-extracted DNA. While common constitutional anomalies were absent in constitutional DNA, several shared CNVs were identified in tumor DNA. The majority of the somatic mutations was shared in 2p22.3, 2q24.3 and 14q11.2. In the single chondrosarcoma studied, large somatic amplifications and/or deletions were observed in chromosomes 3, 6, 9, 10, 12, 13 and 19. Some of these have been reported in other chondrosarcomas using their etiopathogenic, but it remains unknown. NO LOH/UPD was observed in any Maffucci tissue. Our findings identify frequent somatic chromosomal rearrangements, implicating 2p22.3, 2q24.3 and 14q11.2 in the formation of enchondromas and spindle cell hemangiomas in Maffucci syndrome. WES are performed on MS tissues and blood. These alterations and variations are in validation process to identify Maffucci causative mutations.
3435W
A survey of HRAS mutations in a large cohort of patients with Costello Syndrome reveals a predisposition for embryonal rhabdomyosarcoma with paternal uniparental disomy. K.M. Robbins1,2, D.L. Stables1, A. Sadreameli3, J. Hobrook3, S. McCahan1, R. Saha1,2,4, K.W. Grigsby5, K. Sol-Church1.1) Center for Pediatric Research, Al duPont Hospital for Children, Wilmington, DE; 2) Medical Genetics, Al duPont Hospital for Children, Wilmington, DE; 3) University of Delaware Biological Sciences, Newark, DE; 4) Delaware State University Biological Sciences, Dover, DE.

Costello Syndrome (CS) is a rare condition resulting in failure to thrive, intellectual disabilities, short stature, coarse facial features, skeletal abnormalities, and congenital heart disease, caused by heterozygous mutation in the HRAS oncogene. Though most mutations are de novo, originating in the paternal germline, we identified a few cases of somatic mosaicism. CS patients have an increased risk for malignancies, specifically embryonal rhabdomyosarcoma (ERMS). In the general population, non-syndromic (NS) ERMS, is a relatively common pediatric cancer. It is rarely associated with HRAS mutations, but is characterized at the molecular level by loss of heterozygosity (LOH) at 11p15.5. The goal of this study was to molecularly evaluate our cohort, identify dysregulated pathways in CS skin fibroblasts and identify genetic alterations associated with ERMS in CS patients and in NS ERMS. Parental origin of HRAS germline mutations was established using either restriction enzyme digestion or allele specific PCR amplification and sequencing. Differential gene expression analysis was performed with Affymetrix Human Gene 1.0 ST arrays, 6 CS patient derived fibroblast lines and 6 control fibroblast lines. Characterization of the genetic lesions in ERMS was performed using chromosome 11 STR markers, fluorescence in situ hybridization using centromere 11 probe, as well as Affymetrix CytoScan arrays on CS and NS ERMS tumors and cell lines. We molecularly identified HRAS mutations in 104 CS patients. There were 63 cases of paternally derived mutations, 3 maternally derived and the remainder was non-informative. Eleven unrelated patients developed ERMS, and a few relapsed due to residual disease or de novo tumor formation. Nine ERMS tumors from 6 unrelated CS patients were available for molecular evaluation. All patients carried germline paternally derived p.G12S (4) or p.G12A (2) mutations in HRAS. We established a pure ERMS primary cell line using a fresh tumor sample from a patient with a p.G12A germline mutation. Loss of imprinting of H19 and CDKN1C at 11p15.5 was identified, and additional functional analysis and drug screens are currently ongoing. Using CS and NS ERMS tumors, complete LOH was observed along chromosome 11 in CS samples, paternal uniparental disomy is observed in all but one recurrent primary tumor. This molecular characterization will aid in developing new treatments for ERMS.

3436T

It has been estimated that 10% of women diagnosed with breast cancer have a hereditary form of the disease, primarily caused by mutations in BRCA1 or BRCA2, the most prevalent and high susceptibility genes for hereditary breast and ovarian cancer. Worldwide, it is common practice to offer high-risk patients genetic counseling and DNA testing. Even though, molecular approach of these genes is not a common practice in Latin American public institutions because the cost and time consuming of the methods used. We proposed the implementation of massive pyrosequencing for screening of BRCA mutations in patients with suspected hereditary breast/ovarian cancer. Therefore, we used this technology for evaluation of germline mutations in the entire exonic and splice site regions of BRCA1 and BRCA2 in 163 patients with breast and ovary cancer and with familial history of breast cancer or with clinical features suggestive for BRCA mutations. In a first step, the method was evaluated with positive and negative controls with wide and high variability of index cancer genotypes. As a result, we used 80-150 reads per sequence and identified all the BRCA pathogenic mutations; the negative controls did not show deleterious variants, confirming the suitability of their use. Then, we used in the 163 patients previously described. We found 157 (95.7%) patients were carriers of familial mutation and 26 (15.66%) patients were carriers of a widespread founder mutation, p.R337H, detected in 0.3% of the general population. The frequency of BRCA mutations in these patients was 17%. Direct molecular study (PCR and Sanger sequencing) was used to confirm the presence of the putative mutation and in all cases, we discovered a 40% of missense variants; most of them had been reported as variants of unknown clinical significance (VUS). We analyzed VUS according to a multifactorial model. Our results support the suitability of the use of this method in other public institutions, in order to achieve a personalized medicine.

3437T
Clinical characteristics and genotype-phenotype correlations in a large cohort of Brazilian Li-Fraumeni syndrome patients. M. Achatzi1,2, C. Sagne3, P. Ashton-Prolla1,5, V. Márco5,6, A. Nobrega1, J. Hall1, P. Hainaut7. 1) Oncogenetics, A.C. Camargo Cancer Center, Sao Paulo, Sao Paulo, Brazil; 2) National Institute of Science and Technology in Oncogenes (INCTO), Sao Paulo, Brazil; 3) INSEM U612, Centre Universitaire, Orsay, France; 4) Department of Genetics and Molecular Biology, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil; 5) National Institute of Science and Technology in in Populational Medical Genetics (INAGEMP), Porto Alegre, Brazil; 6) Institut Cure, Centre de Recherche, Centre Universitaire, Orsay, France; 7) International Prevention Research Institute, Ecull, France.

Li-Fraumeni syndrome (LFS), an inherited cancer predisposition syndrome, is associated with germ line mutations in TP53. It is characterized by high risk of multiple, early cancers in children and young adults. Carriers of a mutation in DNA-binding domain have a lifetime risk of 90% to develop a cancer. In Brazil, a variant form of LFS is exceedingly frequent due to a widespread founder TP53 mutation, p.R337H, detected in 0.3% of the general population in South and Southeastern Brazil. This mutation occurs in p53 oligomerization domain and its effect is supposed to be dependent upon pH conditions. Individuals with the TP53 p.R337H mutation have lower cumulative lifetime cancer risk (50-65%), with a number of mutation carriers being cancer-free. To identify genetic modifiers of penetrance in TP53 mutation carriers, we analyzed a panel of 87 single-nucleotide polymorphisms in the TP53 locus in 402 members of Brazilian families with breast/ovarian cancer, with or without TP53 mutations. We have identified 2 insertion variants associated with a 12-18 years delay in age at first cancer diagnosis, PIN3 (16bp duplication in intron 3; rs17878362 p=0.082) and DUP3'UTR (6bp duplication in 3'UTR; rs17880560, p =0.067). These SNP are known to regulate p53 pre-mRNA, suggesting that variations in pre-mRNA dynamics may accelerate cancer onset in TP53 mutation carriers. Genotyping these SNP may help in defining personalized follow-up strategies in Brazilian carriers of p.R337H, who have more heterogeneous phenotypes than carriers of ‘classic’ TP53 mutations.
3438W
Exome sequencing identifies potential new candidate genes for unexplained colorectal adenomatous polyposis, S. Aretz1, D. Driehsel2, M. Kerick3, J. Altmueller4, A. Laner5, A. Haaproag1, S. Vogt1, T. Becker2,3, P. Nuenberg1, S. Penner5, E. Holinski-Feder1,6, M.M. Noethen7,10, P. Hoffmann1,10,11, B. Timmermann12, R. Schweiger1, I. Spier1. 1) Institute of Human Genetics, University Hospital Bonn, Bonn, Germany; 2) German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; 3) Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany; 4) Cologne Center for Genomics (CCG), University of Cologne, Germany; 5) Medizinische Klinik und Poliklinik IV, Campus Innens-tadt, Klinikum der Universität München, Germany; 6) Medizinisch Genetisches Zentrum, München, Germany; 7) MVZ Dr. Eberhard & Partner, Dortmund, Germany; 8) Institute of Medical Biometry, Informatics, and Epidemiology, University of Bonn, Germany; 9) Department of Prostate Cancer Research, Institute of Pathology, University Hospital Bonn, Germany; 10) Department of Genomics, Life & Brain Center, University of Bonn, Germany; 11) Division of Medical Genetics, University Hospital Basel, Department of Biomedicine, University of Basel, Switzerland; 12) Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, Berlin, Germany.

Background: In up to 50 % of families with colorectal adenomatous polyposis no germline mutation in the currently known genes APC causing familial adenomatous polyposis (FAP) or MUTYH causing MUTYH-associated polyposis (MAP) can be identified. Methods: To uncover new causative genes, the exomes of eleven unrelated APC and MUTYH mutation negative polyposis patients were sequenced (Illumina HiSeq platform). For data analysis and variant filtering the GATK software 2.1-8, ANNOVAR, and in-house tools were used. Variants and frequent alterations were excluded. Results: In one patient, an APC nonsense mutation in mosaic state (10% of reads, coverage 249) was recognized, which could be confirmed by target sequencing with high coverage (read depths >1000). By Sanger sequencing only a very discreet peak was detectable. In the remaining 10 cases, 66 genes were affected by biallelic truncating variants (recessive model) in at least one patient and 63 genes were affected by truncating heterozygous variants (dominant model) in at least two patients. After detailed visual inspection of the variants in a read browser (Integrative Genomics Viewer) to exclude obvious sequencing artifacts, and data mining according to functions and pathways, six genes of high interest remained (two for the recessive, four for the dominant approach), some of which are involved in cell adhesion, proliferation, or recombination repair. By Sanger sequencing one of the recessive genes (ZSWIM7) and three of the dominant genes could be confirmed (DSC2, HEATRSA and PDE4DIP). Two mutations were artifacts which could not be validated (heterozygous intronic mutation, dominant model) or presented in heterozygous rather than homozygous state (recessive model); the coverage was only 7. Subsequently, one of the two different HEATRSA mutations was also identified in the unaffected mother (no symptoms at 79 years of age, but no colonoscopy). From the other patients no parental blood samples were available. Conclusions: Using exome sequencing we identified new potentially causative genes for adenomatous polyposis. The clinical relevance of the genes is presently clarified in a validation sample of 200 polyposis patients. The study was supported by the German Cancer Aid and BONFOR programme of the University of Bonn.

3439T
Prevalence of BRCA1 mutations in hereditary breast/ovarian cancer families and sporadic triple negative breast cancer patients from Algeria. F. Cherbal1, R. Bakour2, W. Abdou3, C. Mahemmat1, K. Gassi1, H. Gaceb1, K. Boualg2, N. Kanoun-Zitouni3, W. Benbrahim1. 1) Unit of Genetics, LBCM, Faculty of Biological Sciences, USTHB, Algiers, Algeria; 2) Anti Cancer Center, Bida, Algeria; 3) Central Hospital of Algiers, Algiers, Algeria; 4) Medical Oncology Service, Anti Cancer Center, Batna, Algeria.

Background: Breast cancer is the leading cause of cancer death in Algerian women. The present work aimed to establish the frequency of the three most common BRCA1 mutations in Algerian cohort of 96 hereditary breast/ovarian cancer families and sporadic TNBC patients. The identification of common mutations in BRCA1 genes may facilitate genetic testing and counseling. The three BRCA1 mutations screened have previously been reported in hereditary breast/ovarian cancer families and sporadic breast cancer patients from Algeria. Methods: 66 hereditary breast/ovarian cancer families and 30 sporadic breast cancer patients (most of them had triple negative status) were screened for the recurrent BRCA1 mutations c.181T>G and c.798_799delTT. The BRCA1 mutation c.83_84delTG has been screened in 51 hereditary breast/ovarian cancer families. The approach used is based in PCR-direct sequencing. Results: The BRCA1 mutation c.83_84delTG detected previously in two unrelated Algerian breast cancer patients has been identified here in a young breast cancer patient with a strong hereditary breast/ovarian cancer history with the frequency of 1.9% (1/51). Interestingly, the BRCA1 mutation c.83_84delTG has been reported one time in BRCA1 database in Caucasian family, could be specific of Algerian population. The c.181T>G mutation, has been detected here in a young bilateral breast cancer patient with a strong breast cancer family history with the frequency of 1.5% (1/66). To date, the Cys61Gly pathogenic variant is one of the most frequent founder mutation identified in Central European population, has been previously reported for the first time in Maghrebian population in two Algerian and Moroccan families.In addition, haplotype analysis of Maghrebian and central European carriers of the BRCA1 mutation c.181T>G will establish if the origin of this mutation in Maghrebian populations is linked to the Vandsals, an east Germanic tribe, who invaded and established a kingdom in North Africa during the antiquity.The BRCA1 mutation c.798_799delTT detected previously was reported 5 times and 83% in 4 families from Algeria,has not been detected in the present study.The sporadic TNBC patients were negative for the three BRCA1 mutations. Conclusions: the screening for BRCA1 and BRCA2 germline mutations in large series of the Algerian breast/ovarian cancer patients/families will allow to know about the frequency, the spectrum and the contribution of the prevalence of BRCA1 genes mutations.
3440F
Combination of founder mutation screening and genomic capture using BROCA yield high rate of loss of function mutations in early onset and familial breast and ovarian cancer in Greece. F. Fostiari, T. Walsh, S. Casader, M.K. Lee, A. Vratimos, G. Fountzilas, I. Konstantopoulou, M.-C. King, D. Yannoukakos. 1) NSCR Demokritos, Athens, Greece; 2) Dept of Medicine (Medical Genetics), University of Washington, Seattle, WA, USA; 3) Aristotle University of Thessaloniki, Thessaloniki, Greece.

The antiquity of the Greeks as a population defined by language and culture, and the complexity of Greek historical demography, present challenges to genetic testing, including genetic testing for predisposition to cancer. On the one hand, the Greek population harbors ancient founder mutations in many genes, including five damaging alleles of BRCA1, of which three are multiple kb genomic deletions. We screened for these founder alleles in patients with breast and ovarian cancer at a young age (<35 years) or with severe family history of breast, ovarian, and/or pancreatic cancer and discovered that 15% (109/754) carry one of the five mutations. On the other hand, many of the patients who were wildtype at these sites have equally young and familial disease. In order to determine the frequency and range of mutations beyond the Greek BRCA1 founder alleles, we carried out additional Sanger sequencing of BRCA1 and BRCA2 and undertook multiplex massively parallel sequencing using BROCA, which captures the entire loci of 30 breast and ovarian cancer genes, detecting all classes of mutations in each gene. Point mutations were validated by diagnostic PCR and Sanger sequencing. Genomic deletions were validated by MLPA. Of 376 patients with breast or ovarian cancer evaluated thus far, 51 patients (14%) carry loss-of-function mutations, other than the BRCA1 founder mutations, in known breast and ovarian cancer genes. Frequencies are: 25 in BRCA1, 10 in BRCA2, 9 in CHEK2, 2 in PALB2, 2 in ATM, 1 in MSH2 (both patients had endometrioid ovarian cancer), and 1 in MLH1 (in a patient with both colon cancer and early onset bilateral breast cancer). Two missense mutations, CHEK2 p.R180Q and CHEK2 p.G167R, which are known to abrogate DNA repair activity, and one genomic deletion, MSH2 del exons 1-7, have been observed in multiple Greek families and appear to be founder alleles. We conclude that among Greek patients with familial and early onset breast cancer the mutation rate in founder alleles is rather high, the mutational spectrum is nonetheless highly heterogeneous with respect to both loci and alleles. Half of mutation carriers are missed by screening only for founder alleles. A population that has a significant number of founder mutations in breast cancer susceptibility genes can still benefit by an approach that detects all classes of mutations in all known breast cancer genes.

3442T
An unusual case of Lynch syndrome - should chromosome analysis be offered routinely in the investigation of this condition? A. Murray, E. Kirk, M. Prothero, D. Barrell, S. Rolleston, S. Palmer-Smith, P. Thompson. All Wales Medical Genetics Service, Institute of Medical Genetics, University Hospital of Wales, Cardiff, United Kingdom.

Lynch syndrome is an inherited cancer predisposition syndrome, usually caused by germline mutations in a component of the DNA mismatch repair (MMR) system. We present an interesting Lynch syndrome family with an unusual mechanism of loss of MMR function. Our consultand, an asymptomatic, 47-year-old woman, was referred with a family history of colorectal, small bowel, and stomach cancer. We initiated tumour testing on an affected relative, whilst referring our consultand for regular colorectal screening. On her first colonoscopy, an adenocarcinoma was found in the transverse colon. Local hospital immunohistochemical testing of this tumour, for MLH1 and MSH2 only, showed loss of MSH2 expression; molecular genetic testing in our laboratory showed microsatellite instability (MSI), with absence of the BRAF V600E mutation. More comprehensive immunohistochemical testing showed loss of expression of both MSH2 and MSH6, but sequence and dosage analysis of the MSH2 and MSH6 genes did not reveal any mutations. Karyotyping showed the presence of a small paracentric inversion of the short arm of chromosome 2, with breakpoints at p21.1 and p22.2. There is one report in the literature of a similar inversion in a patient with Lynch syndrome although breakpoints were not assigned. This result appears to confirm the clinical diagnosis of Lynch syndrome in our family. We are currently undertaking karyotyping of other affected family members, with a view to offering predictive testing to unaffected relatives, to clarify their risks and screening needs.

3443F
Germline BAP1 mutations in uveal melanoma patients with a personal and/or family history of renal cell carcinoma. R. Pilarski, C.M. Cebulla, O. Saqr, J. Raif, J.B. Massengill, F.H. Davidson, M.H. Abdel-Rahman. 1) Dept. of Internal Medicine and James Cancer Hospital; 2) Dept. of Ophthalmology, Ohio State Univ, Columbus, OH.

Background: Germline mutations in the BAP1 gene have recently been shown by us and others to cause uveal melanoma (UM), mesothelioma, cutaneous melanoma, and possibly other cancers. Noting the occurrence of renal cell carcinoma (RCC) in several families, we investigated the frequency of germline mutations in BAP1 in patients with a personal history of UM and also a personal and/or family history of RCC. Methodology: Six patients were included in the study, including five females and one male. Average age was 52.6 years (range 41 to 72 years). Germline mutations in BAP1 and VHL were assessed in peripheral blood DNA by direct sequencing. In one patient with both UM and RCC somatic mutation in VHL was assessed in both the ocular and renal tumors. Results: We identified germline pathogenic mutations in BAP1 in 2/6 patients. Family histories of mesothelioma and other cancers including lung, colon, breast and pancreatic carcinomas were reported in these two families. No germline VHL mutation was identified; however, a somatic mutation in VHL was detected in the RCC tumor tissue but not in the UM tumor tissue from a patient with both primary tumors. Conclusion: BAP1 is a candidate gene in patients with a personal history of UM and a personal or family history of renal cell carcinoma.

3441W

A 41 year old woman was incidentally found from amniocentesis to have a balanced reciprocal translocation: 46,XX,t(11:12)(p15.3;q14.2). This woman has significant bilateral renal angiomyolipomas, and no other features to suggest a diagnosis of tuberous sclerosis on extensive investigation. The woman's 2-year-old daughter is a carrier of the same translocation. Her parents are not available for investigation. Whole genome sequencing was performed on genomic DNA isolated from blood using the Illumina Hi-Seq 2000. Three lanes of data were generated and a de novo assembly was conducted with ABYSS for chromosomes 11 and 12. The breakpoint was identified in 2 reads and 5 reads from each translocated chromosome. It disrupted NUP98, located at 11p15.4, in intron 1, prior to the translation start site in exon 2. The breakpoint on chromosome 12 was located a gene desert. Nuclear pore complex protein Nup98-Nup96 is encoded by NUP98, a gene recurrently involved in translocations, being fused to an activating partner, associated with blood cancer. Somatic loss of function mutations appear in NUP98 in a variety of cancers. Mice haplo-insufficient for NUP98 have decreased nuclear pore density. This is the first report of a germline translocation involving NUP98 to our knowledge.
3444W
Combined effect of germline mutations in MUTYH and DNA mismatch repair genes on risk of colorectal cancer. A.K. Win1, J.P. Young2, D.D. Buchanan3, S.P. Cleeland4, H. Kim3, J.C. Dowdy1, R.J. Maclnnis1,5, T. Burnett6, L. Le Marchand3, P.A. Newcomb1, R.W. Haile7, N.M. Lindor8, J.L. Hopper2, S. Gallinger3,4, M.A. Jenkins1. 1) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The Univ Melbourne, Carlton, VIC, Australia; 2) Cancer Genetics Group, Queensland Institute of Medical Research, Clive Berghofer Cancer Research Centre, Herston, Queensland, Australia; 3) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Cancer Care Ontario, Toronto, Ontario, Canada; 5) Vancouver General Hospital, Vancouver, British Columbia, Canada; 6) University of Hawaii Cancer Center, Honolulu, Hawaii, USA; 7) Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 8) Department of Medicine, Division of General Internal Medicine, University of Washington, Seattle, WA; 9) Department of Health Science Research, Mayo Clinic Arizona, Scottsdale, Arizona, USA.

Background Identifying modifiers of cancer risks is important for understanding carcinogenesis in carriers of MUTYH mutations and for genetic counselling, screening and risk-reduction strategies. As the MUTYH protein interacts with the DNA mismatch repair (MMR) system, we hypothesized that the combination of a monoallelic MUTYH mutation with an MMR gene mutation increases colorectal cancer (CRC) risk. It has been controversial that MMR-gene mutations, especially MSH6, are more frequent in CRC-affected carriers of monoallelic MUTYH mutation than CRC-unaffected carriers of monoallelic MUTYH mutation. Methods Using the data from the Colon Cancer Family Registry Cohort and Cox proportional hazards regression weighted to correct for the method of ascertainment, we investigated (i) the frequency of MMR gene mutations with an MMR-gene specific MUTYH mutation carriers, and (ii) the effect of monoallelic MUTYH mutations on CRC risk for MMR gene mutation carriers. Results In a cohort of monoallelic MUTYH mutation carriers, we observed that frequency of MMR gene mutations was higher in those with a germline mutation in the RING domain of the MUTYH protein compared to those without a mutation in the RING domain (p=0.02). CRC risk for monoallelic MUTYH mutation carriers with an MMR gene mutation was higher than those without an MMR gene mutation (hazard ratio [HR] 13.2, 95% confidence interval [9.2, 18.4], p<0.001). In a cohort of MMR-gene mutation carriers, we observed that frequency of monoallelic MUTYH mutation was not different between CRC-affected (8/411, 2.0%) and unaffected carriers (2/68, 2.9%) (p=0.64). There was no evidence of difference in CRC risk for MMR mutation carriers with or without a monoallelic MUTYH mutation (HR 0.65, 95% CI 0.24-1.76, p=0.40). Conclusion In this study, there was no evidence of modifying effect of monoallelic MUTYH mutation on CRC risk for MMR gene mutation carriers whereas MMR gene mutation increased risk of CRC for monoallelic MUTYH mutation carriers.

3445T
Towards scoring all 35,397 possible missense variants of BRCA1 for activity. L.M. Stantia1, J.O. Kitzman3, J.G. Chui1, J.D. Pervin2, J. Shendure1, S. Fields2. 1) Departments of Pathology and Genetics, University of Washington, Seattle, WA; 2) Department of Biomedical Informatics, The Ohio State University Comprehensive Cancer, Columbus, OH; 3) Howard Hughes Medical Institute.

BRCA1 is a breast and ovarian cancer-specific tumor suppressor gene that has been subject to much diagnostic sequencing. Multiple cancer-predisposing mutations have been identified along with >500 missense variants classified as Variants of Uncertain Significance or VUS. The BRCA1 protein consists of 3,445T mutations and FAP/AFAP. The majority of patients with a mutation in the three regions associated with AFAP were over the age of 40 at the time of testing (72.7%), while the majority of patients with mutations in the remainder of the gene were aged 20-40 at the time of testing (51.1%). This data suggests that there appears to be a trend of less severe polyposis and later age of testing among patients with a mutation in the three main regions of APC associated with AFAP. However, a substantial number of patients were found to have over 100 colorectal adenomas regardless of the region the mutation was found, suggesting that there is no definitive genotype/phenotype correlation between APC mutations and FAP/AFAP.

3446F
Presentation of patients with mutations in the APC regions associated with AFAP. P. Kaushik, K. Moyes, C. Arnell, M. Landon, R. Wenstrup. Medical Services, Myriad Genetic Laboratories, Inc., Salt Lake City, UT.

Familial Adenomatous Polyposis (FAP) has been characterized by the development of 100s to 1000s of colorectal adenomas, while Attenuated Familial Adenomatous Polyposis (AFAP) has been characterized by the development of less than 100 colorectal adenomas. Both FAP and AFAP are associated with mutations in APC. Studies looking into the phenotype/genotype correlation associated with the APC gene have highlighted specific and consistent features that lead to the differential diagnosis. Three significant areas are the 3’ end, 5’ end, and exon 9. This study aims to describe the polyph trend of patients with mutations in these specific areas compared to mutations in the remainder of the APC gene. A retrospective analysis was performed on patients who underwent full APC gene analysis or targeted APC mutation analysis between January 1st 2006 and May 22nd 2013. Demographics and clinical characteristics were reviewed based on data reported on test request forms submitted to a commercial testing laboratory. Mutational testing results were also reviewed. The regions of the APC gene considered to be associated with AFAP were codons 1-157 (5’ end), 312-438 (exon 9), and 1595-2855 (3’ end). A total of 1534 individuals with mutations in the APC genes were included in this study. APC mutations in the three gene regions associated with AFAP were identified in 461 of 1534 patients (30.1%). Of these, 303 (65.7%) were reported to have developed less than 100 colorectal adenomas. A total of 1073 patients (69.9%) were found to carry an APC mutation in the remainder of the gene not associated with AFAP. Of these, 286 (25%) were reported to have developed less than 100 colorectal adenomas. APC mutations that were not classified as AFAP were identified in the three regions associated with AFAP over the age of 40 at the time of testing (72.7%), while the majority of patients with mutations in the remainder of the APC gene were aged 20-40 at the time of testing (51.1%). This data suggests that there appears to be a trend of less severe polyposis and later age of testing among patients with a mutation in the three main regions of APC associated with AFAP. However, a substantial number of patients were found to have over 100 colorectal adenomas regardless of the region the mutation was found, suggesting that there is no definitive genotype/phenotype correlation between APC mutations and FAP/AFAP.

3447W
High sensitive detection of colorectal cancer mutations using third generation sequencing. G. Rosso1, A. Patrignani2, P. Poveda1, F. Hoehn2, R. Schlappbach1, A. Garvin2. 1) Functional Genomics Center Zurich - ETH/ UZH, Zurich, Switzerland; 2) Droplet Diagnostic SAS, France.

Colorectal cancer (CRC) is a major cause of cancer mortality. Unlike methylation, which occurs to some extent in normal tissue, mutations in the oncogenes and tumor suppressor genes that drive CRC are highly specific for tumor tissue. Mutations in patient samples can be found at many locations in these genes, therefore a scanning method for mutation detection is needed for high sensitivity. Objectives: To determine the sensitivity and specificity of single molecule, third generation sequencing for an assay that detects rare mutations in the genes that drive CRC. Methods: We performed three experiments in which mutant DNA from CRC derived cell lines and mixtures of mutant and wild type DNA were PCR amplified and sequenced on the PacBio RS. High accuracy reads on short amplified fragments were obtained using circular consensus sequencing (CCS) to correct random errors. The assay consists of 15 amplicons covering regions of five genes mutated in CRC (APC, MSS, MSH6, MUTYH, MLH1). These regions are likely to lead to mutations. The amplicons were designed at the same size to ensure uniform coverage and therefore uniform power of detecting mutations. The length of the total test sequence is about 5,000 bp. The results from the PacBio RS instrument were directly compared with Illumina MiSeq on a sample of amplified DNA containing three mutations (two substitutions and one, one base, deletion) present at the 1.5% level. Results: The quality of the CCS-called bases is orders of magnitudes higher than those of the Illumina sequences. Using DNA with mutations present at 0.5% the quality CRC mutations in 15 short amplicons. A variety of mutation frequencies and FAP/AFAP. The majority of patients with mutations in the remainder of the gene not associated with AFAP were identified in 461 of 1534 patients (30.1%). Of these, 303 (65.7%) were reported to have developed less than 100 colorectal adenomas. A total of 1073 patients (69.9%) were found to carry an APC mutation in the remainder of the gene not associated with AFAP. Of these, 286 (25%) were reported to have developed less than 100 colorectal adenomas. APC mutations in the three gene regions associated with AFAP were identified in the three regions associated with AFAP over the age of 40 at the time of testing (72.7%), while the majority of patients with mutations in the remainder of the APC gene were aged 20-40 at the time of testing (51.1%). This data suggests that there appears to be a trend of less severe polyposis and later age of testing among patients with a mutation in the three main regions of APC associated with AFAP. However, a substantial number of patients were found to have over 100 colorectal adenomas regardless of the region the mutation was found, suggesting that there is no definitive genotype/phenotype correlation between APC mutations and FAP/AFAP.
1) Centre for Human Genetics, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 2) Department of Oncology, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 3) Department of Histopathology, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 4) Department of Urology, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 5) Department of Human Genetics, University of Health Sciences (UHS), Lahore, Pakistan.

Abstract: Background: Renal cell carcinoma (RCC) is the most frequent form of kidney cancer in adults and somatic mutations that inactivate the von Hippel-Lindau (VHL) gene are the most common cause of RCC. VHL gene alterations have been associated with clear cell subtype of RCC (ccRCCs), though there have also been reports on other types of RCCs, albeit at lower frequencies. Huge variations have been reported in the frequencies of mutations in the VHL gene in sporadic RCCs from different parts of the world ranging from 45% to 71%. However, there is still no data available on the involvement of the VHL locus in RCC cases from South Asian region. This report presents, for the first time, the genetic and epigenetic changes in a cohort of 300 renal cell carcinoma patients from Pakistan. Methods: To identify mutations in the VHL gene, direct DNA sequencing was carried out and epigenetic silencing was investigated by using methylation-specific polymerase chain reaction. Results: Our data showed molecular alterations in the VHL gene in 144 (48%) renal cell carcinoma patients. Some mutations were found in 87 (29%) patients and 35 novel mutations were identified. VHL promoter hyper-methylation analysis showed epigenetic changes in 83 (39.7%) out of 209 patients. A total of 105 ccRCC patients (52%) and 22 papillary RCC patients (41%) had a molecular abnormality in the VHL gene. A total of 26 RCC cases (12%) carried both types of molecular alterations i.e. somatic mutations as well as the VHL promoter hypermethylation. Patients who had no evidence of molecular alterations in the VHL gene were significantly younger than patients who carried some molecular change. However, there was no significant correlation of a genetic and/or epigenetic change in the VHL gene with other prognostic markers of renal cell carcinomas including tumour grade, size and stage. Molecular alterations in the VHL gene were not restricted to clear-cell renal cell carcinomas (ccRCC).

Conclusions: The comparatively lower frequency of somatic mutations and higher frequency of epigenetic changes in the Pakistani cohort should be considered for developing specifically tailored preventive and therapeutic regimens against non-familial renal cell carcinomas.
Cervical cancer represents the most common cancer death disparity of women in the world, with over 270,000 deaths annually, concentrated in poor, rural, and Indigenous populations. Guatemala and Venezuela are illustrative of this disparity as cervical cancer is the predominant cause of cancer cases and deaths in women. The average age of the subjects is 50, and smoking and oral contraceptive use is low. The affected women have an average of 6 children and adeno- or adenosquamous pathologies represent 25% of the tumors. To determine the genomic changes in invasive tumors, we initiated a prospective collection of invasive cervical cancer tissue. To determine the most frequently mutated genes we sequenced the exomes of 23 tumors in the Ion Torrent Proton and followed up with targeted sequencing of genes by Ampliseq in an additional 80 tumors. The exome sequence revealed potential driver mutations in several known cancer genes such as PTEN, PIK3CA, TSC1, SETD2, HRAS, PTCH, ARID1A, ARID4A and RB1. Targeted resequencing in a larger set of tumors of commonly mutated genes in the COSMIC database (PIK3CA, TP53, STK11, PTEN, KRAS, HRAS, and CDKN2A) revealed 40% mutation in PIK3CA and 11% in PTEN (a negative regulator of the PI kinase pathway). TP53 and STK11 were significantly mutated (10% and 8% of tumors, respectively) but HRAS, KRAS and PIK3CA mutations were partly restricted to a very low level (0.5%). The PIK3CA mutations were concentrated in the known activating sites in the helical domain (E542K and E545K). To determine HPV stains we amplified the HPV L1 gene found detectable in HPV in 96% of samples with HPV16 in 53%, HPV18 in 12%, HPV45 in 10% and HPV26 in 8% of tumors. HPV45 and HPV26 have been previously rarely report. The L1 targets the major HPV capsid protein, which is shared with all HPV types. SNP- plots were applied to identify chromosomal regions of high density, however, in the preliminary results no obvious in intragenic SNP clusters were detected. The results highlight a strategy of an exemplary tumor genome analysis that combines both exome and whole genome sequence information with two different NGS platforms and is one of the first gastric cancer whole genome studies. Using that comprehensive strategy we were able to identify a multitude of novel potentially damaging mutations, which were partially validated in an independent gastric cancer sample cohort of 452 independent samples.
3456W

Massive parallel sequencing of BRCA1 and BRCA2: detection of deleterious mutations and variant of unknown significance in breast cancer patients from Colombia, the new generation of systems biology analysis tools to actively process and evaluate the vast data; how these targets are filtered and screened to identify novel drug targets with the highest probability of success, and the pitfalls identified mining primary, treatment naïve datasets. These methods can then be translated to data released from the ICGC and additional datasets to learn how the disease segmentation for targets identified translate across tumors from diverse populations. 1.http://cancergenome.nih.gov/ 2.http://icgc.org/.

3457T

Understanding the Significance of Individual Tumor Genetic Heterogeneity by Developing Next Generation Genetic Databases as Advanced Analysis Tools. B. Gottlieb1,2,3, L.K. Beitel1,2,3, M. Trifiro1,2,3,4, 1) Dept Cell Gen, Lady Davis Inst Med Res, Montreal, QC, Canada; 2) Segal Cancer Centre, Jewish General Hospital, Montreal, QC, Canada; 3) Dept Human Genetics, McGill University, Montreal, QC, Canada; 4) Dept Medicine, McGill University, Montreal, QC, Canada.

The discovery of intra-tumor genetic heterogeneity in cancer tissues has had a significant effect on cancer ontology by making it even more difficult to assess which gene alterations are drivers and which are passengers within individual patients. To resolve this issue, we have developed a new approach to sequencing that has allowed us to create a distribution profile of mutant variants within tumors. This has allowed us to measure the frequency of distribution of an androgen receptor (AR) repeat length variant associated with breast cancer. Initial data indicate that the relationship of AR mutations and breast cancer is complex and cannot be explained solely by the presence of specific mutations within breast cancer tissues. Understanding intra-tumor genetic heterogeneity and its possible relationship to cancer ontology, is further complicated by the presence of genetic heterogeneity in normal breast tissues as well. Thus, it is becoming clear that most individual cancers will have their own unique genetic profiles. In addition to intra-tissue genetic heterogeneity, the discovery that post-genomic events such as epigenetics can play an important role in determining cancer phenotype, have combined to challenge a number of classic genetic paradigms with regards to the genotype to phenotype relationship. It is the exact nature of this relationship that will need to be resolved to eventually understand carcinogenesis, and we are proposing the creation of next generation genetic databases (NGDBs) as next generation of these NGDBs will incorporate the new genotype to phenotype paradigms, as well as include powerful systems biology analysis tools to actively process and evaluate the vast amounts of both genomic and post-genomic information that will undoubtedly be revealed. The Human Varome Project (HVP) will be an essential element to ensure the effectiveness of NGDBs with its goals of setting up a worldwide network of nation-based genome collection nodes and creating standards for next generation sequencing. Thus, NGDBs, and the HVP can become a new approach to sequencing that has allowed us to create a distribution profile of large genes in a diagnostic setting which is of great importance to meet the increasing expectations of genetic testing. A wider spectrum of at risk women in Colombia will be able to benefit from therapeutic and prophylactic interventions.

3458F

Utilizing Publicly Available NGS Tumor Data to Identify Novel Oncology Targeted Therapies. D. Greenawalt1, J. Bradford1, M. Wappelt1, A. Dunak1, K. Vasudevan2, K. Jacques3, S. Guichard3, J. Dry1. 1) Oncology Bioinformatics, AstraZeneca, Waltham, MA; 2) Oncology Bioinformatics, AstraZeneca, Alderley Park, UK; 3) Cancer Biosciences, AstraZeneca, Waltham MA.

Public tumor consortiums, including the The Cancer Genome Atlas (TCGA1) and International Cancer Genome Consortium (ICGC2) are generating and making publically available a large amount of molecular data on a diverse set of tumor types. Through tumor type specific publications the TCGA is releasing intensive, integrative analysis of specific tumor types and many groups are clamoring to extract novel findings from the data through deep mining and novel integrative analysis. These analyses will lead to an abundance of novel targeted oncology therapies emerging over the next few years. We will present our integrative analysis methods for identifying novel genetic targets from publically available molecular datasets, integrating mutations, gene expression, copy number and fusions from the TCGA data; how these targets are filtered and screened to identify novel drug targets with the highest probability of success, and the pitfalls identified mining primary, treatment naïve datasets. These methods can then be translated to data released from the ICGC and additional datasets to learn how the disease segmentation for targets identified translate across tumors from diverse populations. 1.http://cancergenome.nih.gov/ 2.http://icgc.org/.

3459W

Mutation Status of p53 in Head and Neck Squamous Cell Carcinoma. E. Gunduz1, G. Nasi1, M. Acari1, S. Deede1, K. Erdovan1, C. Moroski Erkül1, M. Gunduz2,3, 1) MEDICAL GENETICS, TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey; 2) DEPARTMENT OF OTOLARYNGOLOGY, HEAD AND NECK SURGERY, TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey.

Head and Neck Squamous Cell Carcinoma (HNSCC) occurs in the Oral Cavity, oropharynx, larynx or hypopharynx and is the sixth most frequent cancer worldwide. Although molecular mechanisms of HNSCC have not yet been revealed completely, p53 alterations are the most frequent event that occurs during carcinogenesis. TP53 is a tumor suppressor protein encoded by TP53 gene, which regulates expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence or DNA repair. Mutations in the p53 gene result in loss of function or decreased levels of gene expression and is associated with a variety of human cancers. Also, p53 inactivation is clinically important because of a strong correlation to an acquired or innate resistance to chemotherapy. Hence, p53 mutation status is one of the most important biomarkers for carcinogenesis and therapy. Our laboratory aimed to identify the mutation status of p53 in HNSCC by full exon sequencing from exon 4 to exon 9 where the most mutations of p53 are located. P53 mutation status was investigated in 31 Head and Neck Squamous cell carcinoma patients and 12 different HNSCC cell lines and these mutations were compared with those in the literature. Our results not only verified previously identified mutations, but also uncovered new mutations that may have functional and clinical importance. Consequently, newly discovered p53 mutations in HNSCC suggest new targets may be important in carcinogenesis, drug resistance and gene therapy.
Next-generation sequencing of paired drug-sensitive and resistant cell lines identifies spectrum of DNA changes associated with drug resistance. P. Jia, H. Jin, C.B. Meador, J. Xia, K. Ohashi, L. Liu, V. Jinawath, N.Jinawath, A.Jinawath, A. Tunteeratam, S. Ngernna, B. Suktitipat. 1) Research Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 2) Department of Biology, Faculty of Science, Mahidol University, Bangkok, Thailand; 3) Illumina, Taiwan; 4) Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 5) Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 6) Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.

Somatic mutations in genes encoding kinases are associated with increased sensitivity of some solid tumors to kinase inhibitors, but patients with metastatic cancer eventually develop disease progression. So far, known resistance mechanisms have been identified such as the second-site mutation, EGFR T790M, amplification of the gene encoding an alternative kinase, MET, and epithelial-mesenchymal transition (EMT). However, the full spectrum of DNA changes associated with EGFR TKI acquired resistance remains unknown. Here, we used next-generation sequencing and bioinformatics analysis to characterize mutational profiles associated with 4 populations of EGFR mutant drug-sensitive cell lines and 5 matched drug-resistant cell lines. We developed a data analysis pipeline to detect SNVs, small insertions and deletions (indels), and CNV changes. By comparing resistant cells with their parental counterparts, we identified 16-89 coding SNVs/indels that were acquired and 1-27 that were lost; few SNVs/indels were found that SNVs/indels tend to occur more frequently in 'constant late' replication timing zones as compared to 'constant early' replication timing zones (chi-squared p-value < 10^-5). An enrichment of SNV frequencies was observed in genomic regions harboring lamina-associated domains compared to the remainder of the nucleus (chi-squared p-value < 10^-5). Surprisingly, we observed a higher burden of CNV changes across all resistant lines, and the one line that had an EMT phenotype displayed significantly higher levels of CNV changes than the other lines with acquired resistance. These results demonstrate a distinct spectrum of drug-resistant phenotypes with different genetic features of the SNVs identified through whole genome sequencing, and how to indirectly assess microsatellite instability (MSI) status were performed. Among the 119 CRCs, the commonly mutated genes KRAS, Braf, and TP53 were mutated in 44%, 40%, and 76%, respectively. Their mutation frequencies in Thai CRCs were in line with those previously reported. Single variant analyses did not identify any variant associated with early-onset CRC. However, using gene-based burden test for rare variants (MAF < 0.02), we identified 10 genes associated with early-onset CRC, namely BRAF, Pdgfra, ATM, Abl1, Egfr, Rb1, Pdk1, Smad4, Kdr, and Id1 (> 0 Gompertz correction). Our results indicate that the development of CRC in non-syndromic early-onset patients may have different underlying genetic alterations than those of average-age-of-onset patients.
Basal breast cancers (BBC) are typically negative for ER, PR and HER2. It is known that the underlying genomic alterations and drug response phenotypes in BBC are different compared to other breast cancer subtypes. There are no validated markers associated with paclitaxel response. We performed the next generation sequencing to investigate biomarkers of paclitaxel drug response in women enrolled in the prospective Breast Cancer Genomic Guided Therapy (BEAUTY) neo-adjuvant clinical study. In this study, women undergo tumor biopsies for NGS at baseline, after 12 weeks of paclitaxel followed by anthracycline-based chemotherapy. Subsets of women with BBC in which both sequencing data along with paclitaxel drug response as measured by MRI were considered in this study. From a cohort of 54 patients, we identified the extremes of phenotype response to paclitaxel in 12 patients with BBC, 5 with evidence for a complete response by MRI and 7 with stable disease (no tumor shrinkage). As expected, BBC showed a high frequency of somatic mutations, 11/12 patients have at least one or more somatic mutations in TP53, PTEN, RB1 and PIK3CA. Several TP53, BRCA1 and MYO3A germline mutations were also identified. In addition, we identified tumor amplification of MYC, PIK3CA, and MYO3A in 7/12 samples. We obtained basal gene expression data for these 12 patients and performed differential expression analysis. The signaling pathways activated and maturity onset diabetes pathway is inhibited in stable group compared to the complete response patients. Significant pathways obtained from DGE analysis were integrated to build a basal transcriptomic landscape (T1) for paclitaxel response. Key nodes in T1 will be further investigated with a variety of genomic features such as splicing, expressed single nucleotide variants, fusion transcripts, and copy number alterations. For this same cohort of BBC, we have also obtained RNA-Seq data after paclitaxel treatment. Hence, we plan to compare and contrast the transcriptomic landscape (T1) with the genomic features obtained from second time point. Integration of such comprehensive OMICS data will help us identify novel molecular targets for BBC treatment.
Exome sequencing identifies different sets of genomic alterations in various subtypes of early-onset breast cancers. C.Y. Lee1, W.H. Kuo2, C.H. Lin1, Y.S. Yu4, P. Lai1, Y.K. Yang3, K.K. Chang2, C.N. Hsiung1, N. Leng1, C.D. Nobuta2, C.H. Chen1, J.Y. Wu1, K.P. Chiu3, A.L. Cheng2, K.K. Chang2, C.Y. Shen5,6, C.J. Chen7. 1) Genomics Research Center, Academia Sinica, Taipei, Taiwan; 2) Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Oncology, National Taiwan University Hospital, Taipei, Taiwan; 4) Molecular and Genomic Epidemiology Research Center, China Medical University Hospital, Taichung, Taiwan; 5) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 6) Department of Medical Research, National Taiwan University Hospital, Taipei, Taiwan; 7) Illumina Inc., Hayward, CA, USA; 8) Department of Surgery, Cheng Ching General Hospital, Taichung, Taiwan; 9) School of Public Health, China Medical University Hospital, Taichung, Taiwan.

Breast cancers in young women manifest differently from those diagnosed at later ages. They are typically aggressive and most patients are candidates for adjuvant therapy. Knowledge of genomic changes is therefore valuable for optimizing treatment strategy. This study identified single nucleotide variations (SNVs) using whole-exome sequencing to provide a comprehensive view of genetic alterations in these patients. Somatic and germline SNVs were investigated in 81 patients diagnosed at 40-years-old or younger. Based on the expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67, tumors were classified into five subtypes: luminal A (ER- and PR-positive, HER2-negative, Ki67 <14%), 34 cases, luminal B-HER2-negative (ER- and/or PR-positive, HER2-negative, Ki67 >14%), 13 cases, luminal B-HER2-positive (ER- and/or PR-positive, HER2-positive, any Ki67, 14 cases), HER2-positive (HER2-positive, HER2-negative, 9 cases), and triple-negative (ER- and PR-negative, HER2-negative, 11 cases). To define which SNVs were functional or related to functional changes, we only included non-synonymous/synonymous changes in regions evolutionally conserved across 44 species. To focus on mutation and not polymorphism, any genetic changes reported in the 1000 Genomes Project and dbSNP were excluded. Nine hundred and thirty-six (877 somatic and 59 germline) SNVs were detected in 176 tumors. The highest-mutation-frequency genes detected in cancers of the same subtype, and tumors of different subtypes identified in cancers of the same subtype, and tumors of different subtypes harbored mutations in the PIK3CA pathway-related genes, but almost all of these (tyrosine kinase receptor genes, G protein-coupled receptor genes and PIK3CA) play an upstream role in the pathway. Mutated genes involved in diverse and unrelated tumorigenic mechanisms were identified in cancers of the same subtype, and tumors of different subtypes were found to harbor genes in the same functional pathways. These findings not only reveal the complexity of the etiology, but also provide valuable clues for developing therapeutic targets and predictive biomarker for early-onset breast cancer.
3469T
Mutational enrichment of cancer-related genes in 11 aggressive prostate cancers. K.J. Lindquist1, R. Kazma2, T.J. Hoffmann1, BA. Rybicki2, A. Levin2, PL. Paris1, JS. Witte1, 1) UCSF, San Francisco, CA; 2) Henry Ford Health System, Detroit, MI.

Prostate cancer is a leading cause of cancer mortality. Many prostate tumors are benign, but some are aggressive and lethal. The mutation profile of aggressive prostate tumors may differ from that of other tumors. To investigate whether somatic mutations in aggressive prostate tumors are more frequent in gene sets previously identified as functionally involved in benign prostate and other cancers, we sequenced the tumor and matched normal tissues of 11 aggressive prostate cancer patients using Complete Genomics’ platform. After removing low-confidence calls, we selected mutations within any gene or gene regulatory region (using data from the Encyclopedia of DNA Elements) in the human genome. Then, we determined if 22 gene sets were associated with mutation rates using a mixed-effects Poisson regression model. The gene sets included 18 cancer-related pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG), 2 differential expression indicators from the Prostate Expression Database, and 2 indicators of cancer drug sensitivity or resistance from the Genomics of Drug Sensitivity in Cancer database. Membership in the KEGG prostate cancer pathway was independently associated with higher mutation rates in our samples (p=0.024), but stronger predictors of high mutation rates were the KEGG transcriptional misregulation pathway (p<0.001), differential expression in response to androgen (p=0.001), and two other KEGG cancer pathways (non-small cell lung and endometrial cancers, p=0.001 for both). Our work paves the way for future studies to link the mutational enrichment of cancer-related gene sets with the aggressiveness of prostate cancers.

3470F
Retrospective analysis of genomic and transcriptional changes in a case of Ewing’s sarcoma tumor progression determined by whole transcriptome and exome semiconductor-based sequencing. G. Meredith1, Y. Sun1, N.S. Hernandez2, M. Taylor1, B. Sanderson1, K. Giorda3, T. Woodburn3, G. Bee1, J. Bishopp3, S. Ghosh3, P. Kapranov4, J. Buckley3, K. Bramliett5, C.P. Reynolds6, T. Triche1. 1) Life Technologis, South San Francisco, CA; 2) Life Technologies, Austin, TX; 3) Life Technologies, Carlsbad, CA; 4) St. Laurent Institute, Providence, RI; 5) Keck School of Medicine of USC, Los Angeles, CA; 6) Texas Tech University Health Sciences Center, Lubbock, TX; 7) The Saban Research Institute Childrens Hospital Los Angeles, Los Angeles, CA.

Ewing’s sarcoma is a pediatric cancer that often presents in the second decade of life and is usually associated with a chromosomal translocation t(11;22)(q24;q12) that results in a EWS/FLI1 gene fusion. Long term survival rates for patients with metastases can be less than 10%. Four independent cell-lines have been established from a patient who succumbed to metastatic disease following relapse after myeloablative chemotherapy. Whole-transcriptome and exome sequencing of normal primary bone marrow-derived stromal fibroblasts (cell-line COG-FB-425), Epstein-Barr Virus (EBV) transformed normal lymphoblasts (cell-line COG-LPS), a pre-therapy primary tumor-derived cell-line (CHLA-9), and a post-chemotherapy metastatic tumor-derived cell-line (CHLA-10), was conducted on an Ion Torrent Proton™ system to profile the differences in gene expression and differences in exonic DNA sequence to characterize the molecular changes associated with primary tumorigenesis and disease persistence after treatment. All cell lines matched by short-tandem repeat analysis. The presence of the EWS/FLI1 fusion gene in the tumor cells was confirmed and the breakpoint determined from normal lymphoblasts (cell-line COG-LPS), a pre-therapy primary tumor-derived cell-line (CHLA-9), and a post-chemotherapy metastatic tumor-derived cell-line (CHLA-10). We used a reverse transcription isoform expression and exon usage between normal, primary tumor, and metastatic tumor cells suggesting an increasing genomic mutational burden in the evolution of the disease, and pointing in particular toward aberrant regulation of RNA-splicing components. One co-expressed, first exon-sharing pair of sense/antisense transcripts corresponded to the gene FEZF1 and transcript FEZF1-AS1 that is unique to the tumor lineage is the subject of further investigation. Taken together, the combination of RNA-seq and exome-sequencing on normal cells and primary vs. post-chemotherapy tumor is providing a broad and deep view of molecular signatures in tumor progression and indicating that a significant role is played by changes in non-coding RNA expression.

3471W
Immunohistochemical analysis of uterine leiomyomas, histopathological uterine leiomyoma subtypes, and uterine leiomyosarcomas. N. Mäkinen1, K. Kämpjärvi1, R. Bitzow2, P. Vaheri2, 1) Department of Medical Genetics, Genome-Scame Biology Research Program, University of Helsinki, Helsinki, Finland; 2) Department of Pathology, The Laboratory of Helsinki University Central Hospital (HUSLAB), Helsinki University Central Hospital and Haartman Institute, University of Helsinki, Helsinki, Finland.

Uterine leiomyomas (ULMs) are benign tumors that arise from smooth muscle cells of the myometrium. Regardless of their benign nature, leiomyomas can cause difficult symptoms, such as excessive bleeding, abdominal pain and discomfort, pregnancy complications, and infertility. The histopathology of ULMs can be divided into common leiomyomas, into various relatively rare subtypes that mimic malignancy in one or more aspects. Cellular, atypical, and mitotically active leiomyomas are examples of these subtypes. Rarely, ULMs may undergo malignant transformation and develop into a leiomyosarcoma. We have previously shown by exome sequencing that MED12 exon 2 is mutated in approximately 70% of ULMs. All the mutations affect an evolutionary conserved region of the MED12 protein. Our subsequent studies have revealed that mitotically active leiomyomas display a MED12 mutation frequency that is not statistically different from common leiomyomas, whereas the rest of the histopathological ULM variants and uterine leiomyosarcomas harbor significantly less MED12 exon 2 mutations than common leiomyomas. The aim of this study is firstly to examine the frequency and respective proportions of known mutations of ULMs, such as mutations in FH and HMGA2, and secondly to scrutinize the possibility that one of the histopathological ULM subtypes would undergo malignant transformation and develop into a leiomyosarcoma by immunohistochemistry. The study material consists of 199 formalin-fixed paraffin-embedded samples including 66 common leiomyomas, 103 histopathological ULM variants and uterine leiomyosarcomas and 30 uterine leiomyomas. We are constructing tissue microarrays of different sample groups for immunohistochemical staining. Antibodies for proteins, such as Ki67 and HIF1α, are utilized to analyze the similarities and differences between the histopathological variants and uterine leiomyosarcomas. It is clinically relevant to recognize possible tumor subtypes in uterine leiomyomas. The molecular genetic changes can be identified on a genome-wide basis, of which some may be involved in the development and progression of tumors, and others in the clinical symptoms. It is also important to understand the mechanisms behind the progression to malignancy, because uterine leiomyosarcomas are usually diagnosed as late as in the surgery.

3472T
LOH Analysis of the ING 3 and 5 genes in Breast Cancer. G. NAS1, E. GUNDUZ1, M. ACAR1, E. UCTEPET1, M. BOZER7, C. DENER2, S. YENIDUNYA2, M. GUNDUZ1,2, E.UCTEPE1, 1) TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey; 2) DEPARTMENT OF TOOTARYNGOLOGY HEAD AND NECK SURGERY, TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey; 3) DEPARTMENT OF GENERAL SURGERY, TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey; 4) DEPARTMENT OF PATHOLOGY, TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey.

The tumor suppressor genes (TSG) ING3 and ING5, the members of inhibitor of growth gene family, are effective in inhibition of cell growth and induction of apoptosis to eliminate cancerous cells. However, in many cancer types, one of the alleles of a TSG is lost through carcinogenesis, while the rest allele is usually inactivated through a process called as loss of heterozygosity (LOH). Previous studies in head and neck cancer revealed that allelic loss and reduced expression is a common pattern of ING gene family members but our study showed that despite the high proportion of LOH in HNSCC, allelic loss of ING3 and ING5 does not commonly occur in breast cancer. 50 paraffin embedded breast cancer tissues were analyzed and the allelic deletion frequency of ING3 and ING5 gene were detected as 12% and 16% respectively. We used to Kaplan-Meier survival analysis to evaluate overall (OS) and disease-free survivals (DFS) in the groups of breast cancer patients with LOH negative and LOH positive ING genes. All the mutations affect an evolutionary conserved region of the MED12 protein. Our subsequent studies have revealed that mitotically active leiomyomas display a MED12 mutation frequency that is not statistically different from common leiomyomas, whereas the rest of the histopathological ULM variants and uterine leiomyosarcomas harbor significantly less MED12 exon 2 mutations than common leiomyomas. The aim of this study is firstly to examine the frequency and respective proportions of known mutations of ULMs, such as mutations in FH and HMGA2, and secondly to scrutinize the possibility that one of the histopathological ULM subtypes would undergo malignant transformation and develop into a leiomyosarcoma by immunohistochemistry. The study material consists of 199 formalin-fixed paraffin-embedded samples including 66 common leiomyomas, 103 histopathological ULM variants and uterine leiomyosarcomas and 30 uterine leiomyomas. We are constructing tissue microarrays of different sample groups for immunohistochemical staining. Antibodies for proteins, such as Ki67 and HIF1α, are utilized to analyze the similarities and differences between the histopathological variants and uterine leiomyosarcomas. It is clinically relevant to recognize possible tumor subtypes in uterine leiomyomas. The molecular genetic changes can be identified on a genome-wide basis, of which some may be involved in the development and progression of tumors, and others in the clinical symptoms. It is also important to understand the mechanisms behind the progression to malignancy, because uterine leiomyosarcomas are usually diagnosed as late as in the surgery.
Characterization of molecular alterations in urologic cancers. M.L. Nickerson1, K.M. Im1, S. Turan1, T. Andresson2, L.E. Moore3, M. Dean1, 1 Cancer & Inflammation Program, National Cancer Institute, Frederick, MD; 2 Laboratory of Proteomics and Analytical Technologies, Advanced Technology Program, SAIC-Frederick, National Cancer Institute, National Institutes of Health, Frederick, MD; 3 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD.

Metastatic cancer is the primary cause of death among patients with urologic and other cancers. Metastatic prostate cancer (mPC) alone is responsible for over 258,000 deaths worldwide each year and is poorly controlled using existing therapies. The molecular chronology from primary to metastatic disease is not known in sufficient detail and detailed molecular markers that differentiate indolent from aggressive subtypes of cancer are needed. We employed exome and targeted sequencing strategies to examine mutated genes in prostate, bladder, and kidney cancers. We present detailed analysis of an index patient with metastatic, castration-resistant prostate cancer where we sequenced the exomes of five metastatic tumors and non-cancer kidney tissue. We observed sequential acquisition of somatic alterations in the primary tumor, including a deletion associated with formation of a TMPRSS2-ERG fusion transcript and a missense alteration in the acetylated lysine binding pocket of a bromodomain in polybromo 1. The majority of somatic alterations, including an alteration of the CpG demethylase, TET2, were not observed in the primary tumor but were observed in all metastatic tumors, indicating mutations likely contributing to metastatic disease may not be detected by analysis of primary tumor tissue. Alterations observed in individual metastatic tumors allowed us to anatomically map the spread of disease. We show TET2 is altered in additional mPC tumors and identify signaling pathways likely affected by TET2 loss. This data is used to discuss observations in urologic cancers that indicate common features of these cancers. We show combinations of rare germline and somatic alterations, and combinations of sequence alterations and loss of heterozygosity differentially target individual cancer genes. Sequential accrual of somatic alterations during disease progression represents a challenge to identify therapies targeting individual cancer genes.

Characterization of somatic alterations in the novel tumor suppressor DEAR1 using ultra-deep targeted next generation sequencing. J. Reuther1, 2, N. Chen1, A. Sahin1, S. Lott1, A. Killary1, 2, 1 Human and Molecular Genetics Program, University of Texas Graduate School for Biomedical Science, Houston, TX; 2 Genetics, MD Anderson Cancer Center, Houston, TX.

Ductal Carcinoma In Situ (DCIS) accounts for 12-24% of all diagnosed breast cancers and is one of the earliest pre-invasive forms of breast cancer. Without treatment, DCIS can progress to invasive disease with recorded frequencies from 14-60%. There is an urgent need to identify prognostic markers for women with a heightened risk of progression from DCIS to invasive breast cancer (IBC) for which more aggressive surveillance and treatment might be warranted, as well as individuals with favorable prognosis, who might be spared rigorous therapeutic regimens and for whom breast conservation therapy might be the preferred surgical option. We previously identified the novel tumor suppressor Ductal Epithelium Associated Ring Chromosome 1 (DEAR1) as a gene, located at 1p35.1, the expression of which is downregulated or lost at the DCIS stage. Our previous work has also shown that loss of expression of DEAR1 by immunohistochemistry significantly predicted local recurrence in early onset breast cancer. Our sequencing efforts, in addition to analyses of the TCGA cohorts, have identified over 30 non-synonymous mutations in DEAR1 in a variety of epithelial cancers including breast, pancreatic and colon cancer, many of which undergo copy number losses within chromosome 1p35. In order to better understand DEAR1’s role in cancer and specifically, its potential role in the progression of DCIS lesions to IBC, our lab is currently undertaking efforts to determine if DEAR1 is mutated in DCIS utilizing the Ion Torrent PGM platform. We are performing targeted ultra-deep sequencing of a 52kb locus containing DEAR1 in both pure DCIS samples and DCIS with associated invasive components, as well as IBC lesions in order to determine the earliest stage in breast cancer in which mutations in DEAR1 can be observed and whether mutations correlate with DCIS that progress to IBC. Targeted deep sequencing has yielded a mean depth of coverage of 9,000x, allowing us to identify both common as well as rare variants. Preliminary results indicate that DEAR1 is mutated in DCIS and IBC and include the identification of novel variants in DCIS as well as missense mutations that have been found previously in IBC. To our knowledge, only few other genes are known to be mutated in DCIS, which suggests the importance of the characterization of this novel tumor suppressor and the determination of its role as a tumor suppressor in breast cancer.
3476F
Individualized analysis of somatic mutations and CNV from exomes obtained from breast cancer core needle biopsies in women with newly diagnosed locally advanced breast cancer. H. Sicotte1, S. Hart1, J.P. Sinwell1, S. Bath1, T.P. Vedel1, K.R. Kalar1, X. Tang2, K.J. Thompson1, D.W. Mahoney1, P. Barman1, J.M. Evans1, C. Wang1, Y.W. Asmann2, J.P.A. Kocher1, T.J. Dockter1, K.N. Jones3, A.L. Connors4, A.M. Moyer1, D.W. Visscher5, V.J. Suman1, J. Eckel Passow1, R. Weinshilboum1, L. Wang1, J.C. Boughey1, M.P. Goetz2. 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Health Sciences Research, Mayo Clinic, Jacksonville, FL; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 4) Department of Surgery, Mayo Clinic, Rochester, MN; 5) Department of Oncology, Mayo Clinic, Rochester, MN.

Introduction: Individualization of cancer therapies based on tumor sequence data is increasing and guidance is needed on what platforms should be used to detect actionable genomic variations or alterations. In the Breast Cancer Genome Guided Therapy Study (BEAUTY), women with locally advanced breast cancer undergo core needle breast biopsy to study genome variations prior to initiating neoadjuvant chemotherapy. These biopsies are profiled using exome sequencing (tumor and germline) and CNV analysis. We will present our methods of analyzing the exome for mutations and somatic copy number variations as well as some cross-platform comparisons. Exome sequencing is rapidly becoming the platform of choice for tumor sequencing because it affords much higher coverage of actionable variants (in coding region and around splice sites of genes) and is cheaper than whole genome sequencing, but it suffers from the lack of accurate tools for calling somatic CNV.

Results: We have developed a somatic CNV calling tool based on the pattern CNV method (submitted), which uses reference samples to learn the pattern and variance of the exome sequencing coverage to better enable somatic CNV calling. This tool is able to detect with high sensitivity EGFR amplifications (FISH validated) in low clonality tumors from core needle biopsies. We will also present the comparison of the efficiency of CNV detection with changes detected using RNASeq. Somatic variant calling in exomes, and in particular low tumor purity samples, is a challenging problem. We are calling somatic variants using multiple somatic calling tools: joint CNV pattern CNV method (submitted), which uses reference samples to learn the pattern CNV method (submitted), which uses reference samples to learn the CNV pattern, a CNV calling tool.

Discussion: Exome sequencing is rapidly becoming the platform of choice for tumor sequencing because it affords much higher coverage of actionable variants (in coding region and around splice sites of genes) and is cheaper than whole genome sequencing, but it suffers from the lack of accurate tools for calling somatic CNV. We are calling somatic variants using multiple somatic calling tools: joint CNV pattern CNV method (submitted), which uses reference samples to learn the CNV pattern, a CNV calling tool.

3477W
Subclonal evolution and genomic drivers of relapse in childhood acute lymphoblastic leukemia. J.F. Spinella1, R. Vidal1, J. Healy1, V. Saillour1, C. Richer1, P. Cassart1, M. Ouimet1, S. Busche2, B. Ge3, T. Pastinen2, D. Sinnett1, 3. 1) Sainte-Justine UHC Research Center, University of Montreal, Montreal, QC, Canada; 2) Department of Human Genetics, McGill University, Montreal, QC; 3) McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada; 4) Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, QC, Canada.

Precursor B-cell acute lymphoblastic leukemia (pre-B ALL) is the most frequent pediatric cancer. Increased understanding of the pathobiology of this disease has led to risk-targeted treatment regimens and increased survival rates. However, close to 20% of patients still do not respond to current treatment protocols. Refractory ALL is the leading cause of death by disease among children. Genetic intratumoral heterogeneity is one of the major mechanisms leading to treatment resistance; whereby clonal evolution and interclonal competition give rise to the accumulation and maintenance of subclonal mutations with a relapse driving potential. However, the clinical significance of these relapse driver mutations and their underlying pediatric ALL which could lead to the development of powerful clinical tools to improve detection, diagnosis and treatment of this childhood cancer.

3479T

Background: Current approaches based on invasive biopsy genetic analysis can fail to capture an accurate picture of the real-time tumor genetic profile due to high tumor heterogeneity and cannot be in practice for serial monitoring of disease progression or acquired resistance. Analysis of circulating tumor nucleic acids (ctDNA), on the other hand, presents a new tool for the monitoring and treatment of cancer. However, due to high-quality false positives in current NGS assays, the majority of studies on ctDNA have been limited to hotspot analyses, amplicon sequencing approaches and typically only involve patients where ctDNA fractions are high (>1-5%).

Methods: We have developed a differentiated sequencing assay, Digital Sequencing Technology (DST) that enables ultra-sensitive and ultra-specific detection of rare genomic abnormalities. Standard NGS workflows are plagued by extremely high noise and distortion in sample-prep and sequencing. DST is able to eliminate the error and distortion created by these processes and produce near-perfect representations of all rare variants. Moreover, our DST workflow enables the vast majority of DNA molecules (even in <10ng input samples) to be converted to sequencing libraries, enabling ultra-sensitive detection.

Results: We first compared the sensitivity and specificity of conventional Illumina SBS sequencing versus DST. We have shown that in sequencing a comprehensive cancer panel of 70kbp in 0.1% cancer cell line titration samples, standard Illumina SBS generates many high-quality false positive variant calls in the range of 0.05-5% (even in <10ng input samples) while DST resulted in highly sensitive and completely error-free variant calls across the entire panel. We then applied DST to rare tumor-derived cell-free DNA on more than 25 cancer patient plasma samples across different cancer types. We investigated the concordance of tumor mutation profiles derived from ctDNA with those derived from matched tumor biopsies. We found higher than 90% concordance between tumor and ctDNA somatic mutation profiles in colorectal, breast, esophagus and melanoma cancers across all patients studied and multiple stages (II-IV).

Conclusion: This work illustrates the potential of ctDNA sequencing for real-time monitoring of disease and it paves the way for ctDNA based personalized treatment.

3479F
Next-generation sequencing of BRCA1 and BRCA2 in Bulgarian breast cancer patients and controls. D.I. Toncheva1,2, L. Balabanski1,3, G. Antov1, I. Dimova1, S. Ivanov1, M. Nacheva4, I. Gavrilov4, D. Neshova1, B. Rukova2, S. Hadzhideкова2, M. Malinov5, 1) Medical Genetics, Medical University-Soﬁa, Sofia, Sofia, Bulgaria; 2) Institute of Genetics, Bulgarian Academy of Sciences, "Acad. Doncho Kostov". Sofia-1113; 3) Department of Medical Genetics, Medical University of Sofia, Zdrave str. 2, 1431 Sofia, Bulgaria; 4) Specialized Hospital for Active Treatment in Oncology, Sofia, 5)Plovdivsko Str. 6, Bulgaria.

Breast cancer is the most common type of cancer in women. Most of its hereditary forms are caused by mutations in the BRCA1 and BRCA2 genes whose main function is DNA repair of the double-stranded ruptures. Genetic testing of women with family history is therefore recommended to determine whether they have hereditary predisposition for this type of cancer. The variants with no clear clinical significance are problematic to make a diagnosis when performing target re-sequencing. After taking the informed consent, DNA samples from 24 patients (average age of 35 ±10) diagnosed with breast cancer and having family history and 71 age matched not affected women were collected. Sequence-targeted BRCA1 and BRCA2 libraries were then prepared using the TrueSeq Custom Amplicon method, and these were sequenced on Illumina MiSeq system. A wide range of variants was found in BRCA1 and BRCA2 genes. In patients' group were found two pathological/presumably pathological variants: mutation in BRCA2 at position chr13:32890665 that affect the first position of the 5' splice region following exon 2 and mutation in BRCA1 at position chr13:32973280, A/-, mononucleotide deletion within 5'-UTR of exon 27; and one pathological variant in BRCA1 at position chr13:32973280, A/-, mononucleotide deletion within 5'-UTR of exon 27; and one pathological variant in BRCA1 at position chr13:32973280, A/-, mononucleotide deletion within 5'-UTR of exon 27; and one pathological variant in BRCA1 at position chr13:32973280, A/-, mononucleotide deletion within 5'-UTR of exon 27; and one pathological variant in BRCA1 at position chr13:32973280, A/-, mononucleotide deletion within 5'-UTR of exon 27; and one pathological variant in BRCA1 at position chr13:32973280, A/-, mononucleotide deletion within 5'-UTR of exon 27;
Recurrence Somatic Mutations in Loss of Heterozygosity Regions of Hepatocellular Carcinoma. S. Tsai, Y. Lin. Molecular and Genomic Medicine, Nat. Health Research Institutes, Zhunan, Miaoli, Taiwan.

BACKGROUND Hepatocellular carcinoma (HCC), the most common liver malignancy, is characterized by frequent loss of heterozygosity (LOH) at multiple chromosomal locations. Genetic alterations in these LOH regions have not been systematically examined. METHODS We investigated 58 target genes in the LOH regions by using Agilent SureSelect enrichment technology in 12 HCC paired samples. Eight candidate genes identified by capture sequencing were further examined by TruSeq custom ampiclon sequencing technology to uncover mutation hotspots in 95 HCC tumors. DNA mass spectrometry was used to validate the findings in independent cohorts. RESULTS Initially we discovered 15 nonsynonymous mutations in 7 genes (CSMD1, DOCK5, PCCA, and PCDH7) and one nonsense mutation in RYR2 gene occurred in at least two patients. CSMD1 is a known tumor suppressor gene and 9/95 HCCs contained a CSMD1 mutation at sites that are critical for protein-protein interactions. Furthermore, a single mutation in PCCA gene encoding propionyl CoA carboxylase occurred in 8/95 HCCs. CONCLUSIONS Focused investigation of selected genes in the LOH regions revealed frequent and recurrent somatic mutations. Genetic alterations discovered by this approach can be used to identify distinct HCC subtypes for developing patient-specific management.

Importance of genetic analysis in the prediction of retinoblastoma in South Indian patients. A. Vanniarajan1, G. Namrata2, K. Usha2, RS. Akram1, K. Thirumalairaj1, J. Jeyaram1, R. Santhis1, R. Kim2, VR. Muthukkaruppan2. 1) Department of Molecular Genetics, Aravind Medical Research Foundation, Madurai, Tamil Nadu, India; 2) Department of Orbit, Oculoplasty and Oncology, Aravind Eye Hospital, Madurai, Tamil Nadu, India. Purpose: India has the highest incidence of retinoblastoma (RB) among the developing countries. Genetic analysis was performed in south Indian RB patients to understand the mutation pattern of RB1 gene in order to predict the risk of RB in siblings and offspring. Methods: The study population included 73 patients of RB who attended Orbit, Oculoplasty and Oncology Department of Aravind Eye Hospital during January to October 2012. A semi log plot was derived to understand the significance of age at diagnosis of RB. Quantitative Multiplex PCR and Exonic sequencing were carried out in tumor samples, wherever available and blood samples of patients and their family members of 73 patients, 31 had bilateral and 42 had unilateral RB. Pedigree showed positive family history in 14 patients (19%) that included 12 bilateral and 2 unilateral RB. Bilateral RB occurred much earlier (9.82 ±11.52 months) than unilateral RB (24.02± 15.11 months), in accordance with Knudson’s hypothesis. Further genetic analysis of tumour samples of 2 unilateral cases showed somatic, nonsense mutations that were present only in tumor but not in blood samples of probands and parents and hence the risk of RB was predicted to be less than 1% in siblings. Among 3 bilateral cases without family history, two patients had deletions of RB1 and one patient had a nonsense mutation in blood, in which is more likely to be inherited to next generation, although parents did not have those mutations. Out of 2 cases with family history the deletions of exons affecting the Pooko domain was observed in both the proband and father. In another family, proband and mother were having a splice site mutation and hence predicted the next sibling had 50% risk of getting RB. When the child was born, clinical examination and genetic testing confirmed RB with same mutation as that of proband. Because of early detection, the child was treated successfully and vision was preserved. Conclusion: Genetic analysis confirmed the Knudson’s two-hit hypothesis in Indian population for the first time. Detection of somatic mutations in sporadic unilateral cases predicted a lesser chance of RB in future generation. The identification of heritable mutations in unilateral cases predicted an increased risk of RB in siblings and next generation, thereby providing early diagnosis and treatment.
3483W
Exome sequencing characterize the somatic mutation spectrum of early serrated lesions in a patient with BRAF negative hyperplastic polyposis syndrome. P. Hoffmann1,2,3, C. Reetz4, C. Strassburg4, H. Froehlich5, P. Nuenberg5, R. Buetnner6, S. Are7. 1) Institute of Human Genetics, University of Bonn, Germany; 2) Department of Genetics, Life & Brain Center, University of Bonn, Germany; 3) Division of Medical Genetics, University Hospital and Department of Biomedicine, University of Basel, Switzerland; 4) Colone Center for Genomics (CCG), University of Cologne, Germany; 5) Department of Internal Medicine I, University of Bonn, Germany; 6) Institute of Pathology, University of Cologne, Germany; 7) German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; 8) Department of Bonn-Aachen International Center for IT (B-IT), Algorithmic Bioinformatics, University of Bonn, Germany.

Background. Hyperplastic polyposis syndrome (HPS) is a very poorly defined colorectal cancer (CRC) predisposition characterized by the occurrence of multiple and/or large serrated lesions throughout the colon. A serrated poly-CRC sequence (serrated pathway) of CRC formation has been postulated, however, to date only few molecular signatures of serrated neoplasia (BRAF, KRAS mutations, CpG Island Methylation, microsatellite instability) were described in a subset of HPS patients and neither the etiology of the syndrome nor the distinct genetic alterations during tumorigenesis have been identified.

Methods. To describe the mutational landscape of serrated polyps and the involved pathways we sequenced the exomes (Illumina HiSeq platform) of 11 early hyperplastic polyps without the BRAF V600E mutation obtained from a 41-year-old female patient with clinically confirmed HPS. For data analysis the VARGANK pipeline of the Cologne Center for Genomics was used. Somatic mutations were identified by comparison with leukocyte DNA and validated by Sanger sequencing.

Results. By analyzing the exome data we initially identified 25 unique somatic alterations in 8/11 serrated tumors. All variants are single basepair substitutions. The predominant mutation type seemed to be missense mutations caused by G>T transversions. However, all but one of the G>T transversions could not be validated by Sanger sequencing, pointing to technical or biological reasons. In seven polyps (ABJB3B, CATSPER8, CCBP2, COL8A1, VGLL2, CALD1, DNA11) are each present in one polyp only. Four mutations occurred in the same single polyp, the other three were identified each in another polyp. No known cancer genes are among the seven candidates, however, some of them are described to be involved in cell adhesion, proliferation, or cell invasion.

Conclusions. Somatic mutations seem to be rare events in early hyperplastic lesions of HPS patients without a BRAF mutation. No frequently affected genes and no enrichment of specific pathways have been observed. Thus, other alterations such as epigenetic changes or variants in regulatory regions might be the major driving force of tumor progression.

The study was supported by the German Cancer Aid (project no. 108421).

3484T

Head and neck parangangiomas (HNPGLs) are rare tumors arising from parasympathetic parangangiolas. 30-45% of HNPGLs are hereditary caused by germline mutations in the von Hippel Lindau gene (VHL). However, the molecular connection between SDH dysfunction and tumor development is still unclear. The most accepted hypothesis proposes a central role for the pseudohypoxic pathway activated by the hypoxia inducible factor (HIF). Paradoxically, we recently showed that activation of HIF in HNPGLs is restricted to a subset of non-SDHx-mutant-HNPGLs. This consisted in accumulation of HIF-1α protein, and over-expression of both, HIF target genes and the HIF-inducible microRNA, miR-210. The present study aimed at unraveling the SDH-independent mechanisms involved in the activation of HIF in HNPGLs. To this end, we analyzed VHL gene in 53 tumors from patients with HNPGLs by direct gene sequencing, Multiplex-Ligation-dependent Probe amplification, qPCR and array CGH. VHL mRNA and protein levels were analyzed by RT-qPCR and immunohistochemistry, respectively. Meta-analysis of the gene expression signature of pseudohypoxic-HNPGLs was performed using Oncomine platform. Meta-analysis of the gene expression signature of pseudohypoxic-HNPGLs revealed that these tumors are highly related to clear cell carcinomas suggesting that HIF/miR-210 pathway is activated via pVHL deregulation in HNPGLs. Accordingly, we identified, for the first time, somatic VHL inactivating mutations (c.482G>A and c.227_229delTCT) in two of four pseudohypoxic-HNPGLs with concomitant LOH in one of them. VHL gene mutations were not found in HNPGLs that lack activation of HIF. However, partial or complete deletion of VHL was found in 32% of non-pseudohypoxic-HNPGLs regardless of the presence or absence of germline SDHx mutations. In addition, low VHL protein levels were detected in 64% tumors in association with decreased VHL mRNA levels. Taken together, our results suggest that the VHL gene is the most frequent target of somatic gene alterations in HNPGLs; and may be an important player in the development of a subset of sporadic HNPGLs via activation of HIF-1α-dependent pathways. Moreover, VHL may also have a role in the development of majority of hereditary and sporadic HNPGLs by activating HIF-independent pathways yet to be identified. This provides the foundation for the development and use of therapeutic approaches that target the pVHL and/or HIF pathway in HNPGLs.

3485F

DNA extraction from formalin-fixed, paraffin-embedded (FFPE) tumor tissues typically results in low quantities of poor-quality, damaged DNA, which presents challenges for sequencing. Here we provide a fast, streamlined FFPE sample-to-answer workflow using probe design software and sample preparation methods optimized to overcome these challenges. DesignStudio® allows for the targeting of up to 1,536 user-defined genomic regions and supports amplicon sizes ideal for characterizing genomic hot-spot regions or sequencing of whole exons with FFPE DNA (amplicons down to 125-bp, with tiling optimums at 175-bp median or higher). The TrueSeq® Custom Amplicon (TSCA) workflow delivers high target enrichment specificity (>80% of reads on target), robust target multiplexing (up to 1,536 amplicons per reaction), and scalable sample indexing (up to 96 libraries). The improved protocol allows lower DNA input (≥100 ng) as compared to previous methods, and supports reduced error rates by means of optimized reagent formulations. Importantly, target representation demonstrates 3x - 47x more read depth in some regions with 80-90% GC. The TSCA assay is highly sensitive, detecting variants down to 2.5% allele frequency using FFPE reference samples verified with digital PCR. As a result of increased MiSeq output, 96 FFPE DNA samples can be targeted with the TrueSeq Amplicon Cancer Panel and sequenced at 1.00xk mean read depth in a single MiSeq run. For germline variant detection, 80 high-quality DNA samples can be targeted with a 1,536-amplicon custom panel and sequenced in a single MiSeq run. In support of a sample-to-answer workflow, the TSCA software analysis tools significantly reduce false-positive variant calls, and enable annotation and filtering of variants. In conclusion, we have developed an integrated FFPE DNA sequencing solution for highly accurate and reproducible detection of somatic variants implicated in cancer biology.
3486W
Alternatively spliced DICER1 transcripts arising from genomic point mutations. M. Wu, L. de Kock, L. Wilkowski, M.R. Fabian, W.D. Foulkes. Medical Genetics, Lady Davis Institute, McGill University, Montreal, Quebec, Canada.

DICER1 is a key endonuclease involved in generating small non-coding RNAs, including microRNAs (miRNAs) and short interfering RNAs (siRNAs). Several germ-line DICER1 mutations have been identified and are associated with a pleiotropic tumour predisposition syndrome. The RNase IIIb domain of DICER1 is responsible for generating 5p miRNAs and is encoded largely by exon 25. We recently identified two tumour-specific somatic single base substitutions (c.5429A>G and c.5438A>G) that result in exon 25 skipping from the DICER1 transcript. We postulate that these point mutations create novel exonic splicing silencer motifs that lead to exon 25 exclusion. After surveying the literature and documenting novel exon 25 mutations from our own studies, we examined whether another 8 other point mutations in exon 25 could also result in its skipping. In total, we found that 4/10 of the exon 25 mutations resulted in either exon skipping or alternative splicing. Here we demonstrate in vitro that the mutation c.5429A>T results in an exon 25 exclusion event while the mutation c.5428G>T creates a cryptic splice site. Importantly, we show that DICER1 transcripts lacking exon 25 or bearing part of exon 25 can be translated. Since exon 25 encodes much of the DICER1 RNase IIIb domain, we planned to determine the impact of these DICER1 mutants on miRNA generation and to compare the functional significance of different mutations in RNase IIIb. This study highlights the importance of examining the effects of genomic mutations at the level of transcript maturation.

3488F
DNA copy number variation and expression of miRNA150 suggestive as prognostic factor for Colorectal cancer. R.V. Andrade1, N. Gasparini2, L. Sakamoto1, G. Pereira2, T. Lins2, 1 Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil; 2 Universidade de Brasília, Brasília, DF, Brazil.

Colorectal cancer (CRC) is the third most common cancer worldwide. It is caused by adenocarcinomas that arise from polyps. The survival rate for CRC is lower as the disease progresses, influencing the prognosis of the patients. Several molecular mechanisms were already described, such as copy number variation (CNV) and miRNA differential expression. MiRNAs can regulate the gene expression by the inhibition of translation or cleaving target miRNAs, and CNVs could regulate gene expression by deletion, translocation, duplication and insertion of these regions. MicroRNA genes embedded in copy number variation can have the expression affected by gene dosage effect. Both molecular mechanisms can be part of cell maintenance, and also modulate the prognosis of cancer. This work aimed to correlate the copy number variation of microRNA genes with clinical features of patients with CRC. Copy number variation and microRNA were evaluated by qPCR using TaqMan assays in normal, tumor and lymph node tissues. The Mantel-Cox test along with the Kaplan-Meier test was used to determine the survival time between the group of diploid (2 copies of microRNA genes) and non-diploid carriers. The expression of miR150 was statistically different between normal and tumor (p<0.001), and between normal and lymph node tissues (p<0.001) showing that the expression of tumor is downregulated in relation to normal tissue, and the expression of lymph node is upregulated in relation to normal tissue. However, there was no significance between survival rates and up or downregulation (p>0.05). For the CNV, the Wilcoxon test did not indicate statistical difference in the gain of CNV between normal and tumor, and with lymph node tissues (p>0.748). However, when groups were defined by diploid and non-diploid individuals it indicated that the CNV-miR150 diploid genotype is associated with lower survival rates (p<0.022). Data from miR150 is associated with lower survival rates. This indicates that CNVs containing genes of variable nature may behave differently to the gene dosage altering the functional expression in a tumor sample. For future experiments it is suggested to evaluate whether or not CNVs in addition to miRNAs and quantify their expression in relation to normal tissue.

3489W
MicroRNA-192 regulates cellular proliferation in medulloblastoma seeding by targeting of dihydrofolate reductase. S.A. Choi1,2, S.Y. Yang1, H.A. Kim1,2, K.C. Wang1,2, J.H. Phi1,2, J.Y. Lee1,2, J.H. Choi1,2, D.Y. Um1,2, Y.J. Moon1,2, P.A. Kwak1,2, E.J. Kwon1,2, S.K. Kim1,2. 1) Division of Pediatric Neurosurgery, Pediatric Clinical Neuroscience Center, Seoul National University Children's Hospital; 2) Adolescent Cancer Center, Seoul National University Cancer Hospital; 3) Division of Neurosurgery, Dongguk University Hospital.

Medulloblastoma (MB), a most common malignant brain cancer in children, frequently disseminates throughout the cerebrospinal fluid. The treatment of leptomeningeal dissemination (seeding) of the MB is challenging and the prognosis is extremely poor. In this study, we profiled miRNA expression and investigated the effect of miR-192 regulated DHFR (Dihydrofolate reductase) expression to get a better understanding of biology in MB seeding. miRNA profiling was carried out using agilent human miRNA microarray kit and qRT-PCR and western blot were performed using tissues and MB cell lines. For the further studies, MB cell lines were transfected miRNAs and investigated the luciferase activity, viability, proliferation assay and cell cycle analysis. The expression level of miR-192 is down-regulated in seeding group and three different MB cell lines compared to non-seeding MB and/or normal cerebellum. MiR-192 directly binds to the 3'-untranslated region (3'UTR) of DHFR mRNA, resulting in over 50% decrease in DHFR protein level in MB cell lines. The overexpressed miR-192 reduced the cell proliferation and induced cell cycle arrest by controlling DHFR target protein. The miR-192 overexpression considerably inhibited vimentin protein level suggesting that it might control the anchorage-dependent cell growth not migration. Our results demonstrate that DHFR is a direct target of miR-192 regulation and overexpression of miR-192 reduced tumor cell proliferation by inhibiting the expression level of the DHFR. Our findings describe a new mechanism for the regulation of DHFR/miR-192 and vimentin link providing a better understanding of signal pathway of MB seeding.
These mutations and suggests the existence of other yet unknown genes that were possibly exposed to asbestos developed MM, a further individual that BAP1 is not expressed in tumor tissue, according with Knudson's two germline BAP1 mutations in sporadic MM. One family carried a new LOF mutation in the BAP1 oncosuppressor gene is responsible for an inherited syndrome with predisposition to MM, uveal and keratinocytic melanoma, renal carcinomas and other malignancies (Testa et al 2011, Popova et al 2013). Predisposition to MM may also occur with an autosomal dominant fashion. Germline loss-of-function (LOF) mutations in the BAP1 gene cause susceptibility for inherited cancer syndrome with predisposition to MM, uveal and keratinocytic melanoma, renal carcinomas and other malignancies (Testa et al 2011, Popova et al 2013). So far germline mutations in BAP1 were identified in 28 families, 9/28 showed multiple cases of MM and 3/28 showed a single case of MM. Germline mutations have been identified also in sporadic MM cases (2/26, Testa et al 2011). In this study, we report the analysis of BAP1 in four multiplex families and in 103 sporadic MM cases with the two aims of further investigating BAP1 related cancer syndrome and of estimating the role of germline BAP1 mutations in sporadic MM. One family carried a new LOF germline mutation (c.46_47insA). By using immunohistochemistry we show that BAP1 is not expressed in tumor tissue, according with Knudson's two hits hypothesis. Interestingly, whereas the three patients (mutation carriers) that were exposed to asbestos developed MM, a further individual (mutation carrier) who was not exposed developed a different tumor type, i.e. mucoepidermoid carcinoma. This suggests that carcinogen type exposure may be important for the cancer type that is developed by mutation carriers. The other families did not show mutations in BAP1. None of the 103 sporadic patients showed mutations in BAP1, with an estimated confidence interval computed using Poisson distribution from 0 to 3.58%. Our data show that germline BAP1 mutations have a very limited role in sporadic MM. On the other hand, our study focused on familial aggregation of MM identifies a new BAP1 mutation, extends the cancer types associated with these mutations and suggests the existence of other yet unknown genes in the pathogenesis of familial MM.
Expression profiling of cofilin-1 in breast cancer cell lines and biopsies. A. Hadjisavvas1, C. Sutton2, S. Shaheed2, P. Loadman2, V. Speirs2, A. Hanby3, K. Kyriacou1. 1) The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 2) Institute of Cancer Therapeutics University of Bradford, Bradford, UK; 3) Leeds Institute of Molecular Medicine, Leeds, UK.

Breast cancer is a significant cause of death since many tumors do not respond to treatment or acquire resistance. Hence there remains the need to identify novel targets for therapy and biomarkers for prediction of response to treatment. A quantitative proteomics study of matched normal and tumor biopsies identified 63 proteins to be significantly increased or decreased in stage-specific cancer. The use of these biomarkers could improve the diagnosis of cancer, while others such as cofilin-1 represent new targets for investigation. Cofilin-1 was subject to a range of analyses to determine its association with breast cancer. Western blotting (WB) was performed on protein extracts from breast cancer cell lines and matched normal and tumor biopsies from Cypriot patients with different stages of the disease. Immunohistochemistry (IHC) was carried out on core biopsies. Multiple reaction monitoring (MRM) mass spectrometry was performed on trypsin-digested protein extracts from biopsies. WB indicated that cofilin-1 was ubiquitously expressed in breast cancer cell lines, representative of Luminal A and B, basal-like, claudin-low and HER2 phenotypes, at higher levels than normal breast cell lines. WB also indicated increased expression of cofilin-1 in invasive carcinoma tissues, compared to matched normal. Patients with ductal carcinoma in situ of fibroadenoma exhibited less clear results, either increasing slightly or remaining unchanged. MRM analysis of three cofilin-1 peptides in tissue extracts of invasive carcinoma patients indicated expression only in tumor. The IHC of core biopsies exhibited strong staining for cofilin-1 in ductal invasive carcinoma, while expression was negative in ductal carcinoma in situ and in benign tissues. Bioinformatics provided further confidence in our findings, highlighting the value of utilising databases for evidence of disease-specific proteins.
Contralateral Mastectomy and Survival after Breast Cancer in BRCA1 and BRCA2 Mutation Carriers. K. Metcalfe1, S. Gerstein1, P. Ghidainan2, H. Lynch3, C. Snyder4, N. Tung3, C. Kim-Sing5, A. Eisen6, W. Foulkes6, B. Rosen8, P. Sun7, S. Nardod1. 1) Faculty Nursing, Univ Toronto, Toronto, ON, Canada; 2) Epidemiology Research Unit, Centre Hospitalier de Universite de Montreal (CHUM), Montreal, Canada; 3) Department of Preventive Medicine and Public Health, University of Medicine, Omaha, NE; 4) Beth Israel Deaconess Medical Center, Boston, MA; 5) BC Cancer Agency, Vancouver, Canada; 6) Toronto Sunnybrook Regional Cancer Center, Toronto, Canada; 7) Program in Cancer Genetics, McGill University, Montreal, Canada; 8) Universite Laval Health Network, Toronto, Canada; 9) Women’s College Research Institute, Toronto, Canada.

Background: Women who carry a mutation in either the BRCA1 or BRCA2 gene face a lifetime risk of breast cancer and once diagnosed, face a high risk of secondary breast cancer. Contralateral mastectomy is an option to reduce the risk of contralateral breast cancer. However, it is unclear if contralateral mastectomy impacts on survival. The objective of the current longitudinal study was to compare the survival rates of women with BRCA-associated breast cancer who did and who did not undergo a contralateral (bilateral) mastectomy. Methods: Patients were 390 women with stage I or II breast cancer and a BRCA1 or BRCA2 mutation who were initially treated with unilateral or bilateral mastectomy. These patients were selected by pedigrees of families who received counseling at one of 12 participating clinical genetics centers. The medical records were re-examined for clinical presentation, medical and surgical treatments and outcome. Patients were followed for up to 20 years from diagnosis. Survival experience was compared for women who did and who did not have a contralateral mastectomy. Results: The probability of breast cancer death at 10 years was 5.9% in women with a contralateral mastectomy and 7.8% in women with breast cancer who did not undergo a contralateral mastectomy (HR = 0.75; 95% CI: 0.51 to 1.1; p<0.05). Conclusions: BRCA-positive women who are treated for breast cancer with bilateral mastectomy are less likely to die from breast cancer than women who are treated with unilateral mastectomy.

Prevalence of Succinate Dehydrogenase-deficient GIST in adults with GIST. I.R Rainville1, E.J. Root2, A.J. Wagner3, S. George4, C.L. Conless5, J.E. Gubler6, J.A. Honnick1. 1) Center for Cancer Genetics and Prevention, Dana-Farber Cancer Inst, Boston, MA; 2) Center for Sarcoma and Bone Oncology, Dana-Farber Cancer Inst, Boston, MA; 3) Department of Pathology, Brigham and Women’s Hospital, Boston, MA; 4) Portland VA Medical Center and Oregon Health and Sciences University Knight Cancer Institute, Portland, OR.

Introduction: GIST is the most common mesenchymal tumor of the GI tract, often associated with somatic gain-of-function mutations in KIT or, less frequently, PDGFRA. Loss of heterozygosity in tumors suggests a somatic phenotype similar to the associated retinoblastoma. Succinate dehydrogenase distinguishes a class of GISTs wild-type for KIT/PDGFRα with distinctive histopathologic and clinical behavior. Germline mutations in SDHX with somatic loss of heterozygosity in tumors characterize the syndrome, and patients are at risk of second primary malignancies. We reviewed 50 randomly selected cases of apparent sporadic GIST, and correlate with loss of expression of SDH subunits by immunohistochemistry (IHC). We are systematically reviewing GIST pathology, family history, and germline status in our multicenter study to estimate the prevalence of SDH-deficiency among adults with GIST. We report the results from our data and Dana-Farber Cancer Institute (DFCI) cohort thus far. Methods: Individuals ≥ age 18 years with history of confirmed GIST were enrolled through the sarcoma clinic at the DFCI. Gastric GISTs were screened for features of SDH deficiency; epithelioid or mixed morphology, multinodular or pleomorphic growth pattern, and lymph node metastasis. Somatic KIT/PDGFRA genotyping was noted when available. Suspected SDH-deficiency was confirmed by IHC for SDH-A/B. Germline testing, if not done previously, was offered to study participants. Results: From March, 2011 to February, 2013, 34 patients with GIST were included with 19 (56%) who had SDH defects on IHC. Of those, 17 (89%) of 19 with GIST, 18 (95%) had IHC positive results in the DFCI sarcoma program. To date, IHC has identified 19 SDH-deficient GISTs (9.9%; CI 5.1% to 12.7%). Age of diagnosis ranged from 15 to 55 years (mean 31.9 years). KIT/PDGFRA mutation screening was available for 7 (37%) of 19 cases, all of which were negative. Seven (37%) of 19 with family history were reviewed were apparently sporadic. Three cases had family history of GIST in a first or second-degree relative. Germline SDHX analysis of 9 cases performed in our study or elsewhere identified mutations in 2 in SDHB and 1 in SDHC. Conclusion: SDH-deficient GIST comprises a significant fraction of all GISTs in this cohort. Family history of GIST or paraganglioma, when present, can inform genetic testing. In the absence of family history of other associated tumors, tumor histopathology with SDHA/ SDHB analysis are of particular relevance for counseling their sisters and daughters. These results are of particular significance for counseling their sisters and daughters. These results are of particular significance for counseling their sisters and daughters.
Long intergenic non-coding RNAs (lincRNAs) have been found to perform various functions in a series of important biological processes. Because the transcription of lincRNAs is tissue and temporal specific, it is important to examine a variety of tissue types as part of the discovery process. Recently, genome-wide transcriptome deep sequencing technology and computational approaches have provided an unprecedented opportunity to analyze such transcripts. In this pilot study, we performed RNA-seq on 40 normal prostate tissue samples to investigate and characterize the entire transcriptome. Regions of normal prostate tissue were histologically confirmed from specimens obtained during radical prostatectomy for prostate cancer (n=38) or cystoprostatectomy in patients with bladder cancer (n=2). Specifically, we focused on the identification of prostate-specific lincRNAs. In total, we identified 22,176 candidate lincRNA transcripts at 16,145 loci. Candidate lincRNA were defined as having evidence from assembly for at least two samples. Overall, there was significant overlap between the lincRNAs identified in this study with lincRNAs annotated in ENCODE V16 (75% overlap at loci level). Importantly, 5544 novel lincRNAs were not previously reported. The prostate derived lincRNAs were further examined for evidence of transcriptional activity using the H3K4me3 and H3K36me3 marks across the entire locus. Of the remaining transcripts, 2797 (27%) overlap a H3K4me3 peak alone (promoter region) and 1221 (12%) overlap a H3K36me3 peak alone (transcribed region). Some future directions to characterize these lincRNA transcripts include differential expression analysis of these lincRNAs between different clinical groups, co-expression analysis of lincRNAs and protein-coding RNAs, and analysis of disease-association between different groups, disease-association analysis of these lincRNAs between different clinical groups, co-expression analysis of lincRNAs and protein-coding RNAs, and analysis of disease-association.
3504W
Whole-Exome Sequencing of Familial Aggregations of Radiation-Associated Meningiomas (RAM). A. Pathak1, R. Bruchim2, D.R. Stewart3, S. Sadetzki1. 1) NIH NCI Division of Cancer Epidemiology and Genetics, Rockville, MD; 2) Cancer & Radiation Epidemiology Unit, Genter Institute, Israel.

Sporadic and radiation-associated meningiomas are rare diseases in the general population, and familial meningioma occurs in (5-10%) of all cases. Our group has previously demonstrated that 11% of families with RAM have two to four first degree relatives who develop meningioma after exposure to radiation in childhood. However, the key genetic drivers of this ~25-fold increased cancer susceptibility have yet to be fully elucidated. We subjected 18 individuals from 6 RAM families to whole-exome sequencing. All individuals in the study were irradiated for lyme capsitis. Five families had 2 affected individuals and 1 unaffected individual. One family had 1 affected individual and 2 unaffected individuals. The study population was derived from the Israeli Tinea Capitis (TC) cohort which included a group of 10,842 individuals treated during the 1950s with radiation therapy for TC. The sequencing revealed 42,860 non-synonymous variants. In our initial bioinformatics analysis, we implemented a custom PERL script to perform automated PubMed queries for the 706 genes of biological interest and the keyword “meningioma.” This resulted in the identification of 16 variants genes associated with meningioma. Further investigation of this literature revealed that a missense mutation in C1D and a stop codon in UBAP1 (a protein involved in the ubiquitinization pathway) may be associated with RAM. Next, we filtered these variants based on disease status and ESP count (<0.5%). In addition to the stop codon in UBAP1, we also identified a stop codon in PTPN7, a protein tyrosine phosphatase. For missense mutations, we further filtered our variants based on having a polyphen-2 score of possibly damaging or probably damaging. We prioritized these variants based on the literature and based on occurrence in multiple families. We found 5 variants in all that occurred in several families. Specifically, we detected two damaging mutations in the RNA polymerase POL2RF and a single damaging variant in two families in a ubiquitinization gene UBQLN3. Further sequencing of these candidate genes in larger validation groups may elucidate the molecular mechanisms underlying the significant genetic predisposition to RAM. The fact that no single locus was found in these families may indicate that the mode of inheritance of familial RAM may be due to the combination in incidence of several low-risk variants rather than a single high-penetrance gene.

3505T
Global differential expressions of isoforms of miRNA in retinoblastoma: correlation with level of expressed 3’-5’exonucleases. A. Gran-guly1, J. Leipizig2, J. Richards-Yutz1, J. Purrazzoellia1. 1) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 2) Bioinformatics Core Center for Biometric Informatics The Children’s Hospital of Philadelphia.

miRNAs are short, 21-24nt, oligonucleotides that function as post-transcriptional regulators of gene expression. Dysregulation of miRNA expression plays an important role in cancer. Retinoblastoma (RB) is the most common intraocular malignancy of childhood. To understand the role of miRNAs in RB, we profiled the miRNA transcriptome of normal neonatal retina, two RB cell lines, WERI and Y79, and three matched retina/RB samples using next-generation sequencing on Illumina platform.

Members of the let-7 family of miRNAs were differentially expressed between normal retina and two RB derived cell lines when sequenced on GAIIx platform. Additionally, pervasive presence of length variations at the 3’-ends of miRNAs, specifically in let-7b, let-7i and miR-181b were observed. The number of reads corresponding to each miRNA was greater than 100 in any one sample. The proportion of canonical to variant miRNAs was statistically different between normal retina, retinoblastoma and two cell lines - WERI and Y-79. The differential expression of EXD2, a 3’-5’exonuclease was statistically different between normal retina, retinoblastoma and two cell lines - WERI and Y-79. The expression of EXD2 correlated with the relative abundance of the canonical/variant miRNA and with the degree of differentiation/aggressive behavior of the retinoblastoma tumors.

3506F
MALAT1 is deregulated and co-expressed with CREBBP in the long non-coding transcriptome of childhood Acute Lymphoblastic Leukemia. R. Vidal1, C. Roger1, J.F. Spinella2, V. Salliguer2, M. Ouimette1, S. Langlois1, P. Cassart1, J. Healy1, E. Bareko1, A. Droit1, D. Sinnett1-2. 1) Centre de recherche, CHU Sainte-Justine, Montreal, QC, Canada; 2) Département de Pédiatrie, Faculté de Médecine, Université de Montréal, Montreal, QC, Canada;

MALAT1 is one of the most highly expressed IncRNAs with the highest expression in all pediatric cancers. MALAT1 is co-expressed with CREBBP in the long non-coding transcriptome of childhood Acute Lymphoblastic Leukemia (ALL). We used the Illumina platform to quantify 46,000 IncRNAs in two cell lines and three matched pairs of matched retina/RB samples. Our results revealed 36,200,000 reads per sample (90% alignment) with median 45,800 reads per isoform. Of these, 65,000 canonical isoforms were predicted with at least 40 reads per isoform. We found a strong correlation between the expression of MALAT1 and CREBBP in childhood ALL, which is consistent with our previous findings in adult patients. In contrast, MALAT1 expression is not significantly correlated with CREBBP in adult ALL. These findings suggest that the long non-coding transcriptome of childhood ALL is distinct from that of adult ALL and may represent a distinct biological signature.

3507W
Prediction of GWAS-identified risk loci in breast cancer and correlated SNPs through mapped epigenetic phenomena. P.S. Rajagopal1, Q. Li2, P. Kraft1, M. Freedman2. 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA.

INTRODUCTION: The majority of loci identified from GWAS are located outside of known protein-coding regions. Projects such as ENCODE and GENCODE have begun epigenetic and non-coding RNA annotation of these regions. These databases have been used to associate GWAS findings en masse with putative function. However, this has been minimally explored for individual diseases, individual SNPs or predictive ability. Using breast cancer as an example, we correlated GWAS-identified risk loci with functional information from these databases to create a model that predicts a SNP’s likely association with disease. METHODS: 76 index SNPs and 946 proxy SNPs (r²>0.9, based on the Cancer Genome Atlas ALL cohort) were selected based on the COGS breast cancer GWAS (p<10⁻⁷). These were matched to randomly selected non-GWAS SNPs by chromosome, minor allele frequency (MAF), r² and number of proxy SNPs. Information was collected on GRCh37:12 position, MAF and number of index SNPs, total number of proxy SNPs, and number of GWAS SNPs in the region. We evaluated these elements for enrichment among GWAS-associated SNPs compared to the randomly selected SNPs. For each index SNP, binary counts were summed across all proxies such that values per SNP ranged from 0 to 1 and the total number of proxies for that SNP (sum). Variables were also created for any proxy having a trait (binary) and average across all proxies (average). A conditional logistic regression model was developed using forward selection (p<0.05). RESULTS: Our final model included: RegulomeDB score ≤4 (binary) (OR=4.79, 95% CI: 52.15-17.13, p=1.2E⁻⁸), RegulomeDB score >4 (binary) (OR=1.13, 95% CI: 0.99-1.29), MCF-7 RNA polymerase II binding (binary) (OR=9.02, 0.97-84.04) and SNP presence in transcripts (average) (OR=2.51, 95% CI: 0.95-6.68). MHC Nase I hypersensitivity (average) correlated closely with RegulomeDB score. We lacked enough data to evaluate the relative role of all individual RNA types, H3K4me3 or CTCF. CONCLUSIONS: We have demonstrated the feasibility of deriving a SNP-based predictor of association with disease based on genomic and epigenetic features enriched in known risk loci vs. random loci. This method synthesizes readily accessible information to bring GWAS findings to the post-GWAS era.