ASHG 2022 Plenary Abstracts

As of September 30, 2022

Instructions:

1. View the Table of Contents on the next page.
    a. The Table of Contents is sorted chronologically by Program Number.
2. Click on the Program Number/Abstract Title of the abstract you would like to view and you will be taken to the page the abstract is on.
3. Details included with each abstract are: the session number and name, the presentation location, session time, abstract title, author block, and abstract body.

Table of Contents

S03. Featured Plenary Abstract Session I

Location: Conv Ctr/West Hall A/West Building

Session Time: Tuesday, October 25, 2022, 2:15 pm - 4:00 pm

ProgNbr 014: Deep learning to understand the genetic architecture of the human skeletal form

Authors:

E. Kun[1], E. M. Javan[1], O. Smith[1], F. Gulamali[2], B. I. Flynn[1], Z. D. Trutner[1,3], K. Vajrala[1], M. Sohail[4], P. Jayakumar[1,3], T. Singh[5,6], V. M. Narasimhan[1,3]; [1]The Univ. of Texas at Austin, Austin, TX, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Dell Med. Sch., Austin, TX, [4]UNAM, Cuernavaca, Mexico, [5]Columbia Univ., New York, NY, [6]New York Genome Ctr., New York, NY

Abstract:

The human skeletal form underlies our ability to walk on two legs, but despite large sample size genetic studies for human height, less research has focused on limb lengths and overall skeletal proportions. To investigate the genetic architecture of the human skeletal form, we applied a deep learning model to 31,221 whole body dual-energy x-ray absorptiometry (DXA) images from the UK Biobank (UKB) to extract 23 different image derived phenotypes by measuring all long bone lengths as well as widths of the hip and shoulder and then converting these measurements to skeletal proportions while controlling for height. We verified the accuracy of our phenotyping through checks of left-right symmetry and duplicate patient cohort images taken across two time points, which all had correlation rates of greater than 90%. We also compared our deep learning derived measurements to human annotated measurements and found a difference of less than 3 pixels (<1 cm) for each phenotype. All skeletal proportions were found to be highly heritable (~40-50%) and a genome wide association study (GWAS) of these traits identified a total of 183 loci that were height independent. These loci are associated with genes active in skeletal morphogenesis and development, rare human skeletal diseases, and abnormal mouse skeletal phenotypes. We then conducted phenotypic associations, genetic correlations, and polygenic risk score analyses of skeletal proportions with a range of musculoskeletal disorders and discovered significant associations particularly with osteoarthritis, a leading cause of disability in the United States. To connect the genetics of skeletal lengths to growth plate biology, we co-analyzed our GWAS results with gene expression from three dissected layers of murine newborn tibial growth plate and discovered significant enrichment in the round layer of the bone. Furthermore, as skeletal anatomical traits are strikingly different between humans and the great apes, we performed evolutionary analysis. In contrast to GWAS loci from cardiovascular, auto-immune, metabolic diseases as well as a range of other quantitative traits, loci identified by our GWAS of skeletal proportions showed enrichment in Human Accelerated Regions as well as in regulatory elements of genes that are differentially expressed through development between humans and the great apes. Taken together, our work validates the use of deep learning models to extract phenotypes from medical images, identifies genetic variants that affect the overall human skeletal form - an important part of our biomechanical shift to bipedalism - and links an evolutionarily novel aspect of human anatomy to its pathogenesis.

S03. Featured Plenary Abstract Session I

Location: Conv Ctr/West Hall A/West Building

Session Time: Tuesday, October 25, 2022, 2:15 pm - 4:00 pm

ProgNbr 015: Leveraging sequencing data in the UK Biobank to detect parent-of-origin effects in 80,821 individuals

Authors:

R. J. Hofmeister, S. Rubinacci, D. Ribeiro, O. Delaneau; Univ. of Lausanne, Lausanne, Switzerland

Abstract:

Identical genetic variations can have different phenotypic effects depending on their parent-of-origin (PofO). Yet, studies discovering PofO effects are limited in sample size, as they rely on parental genomes or known genealogies. While modern biobanks gather genotypes and phenotypes across hundreds of thousands of individuals, only limited genealogies are available, preventing the study of PofO effects in such datasets.
We present a novel probabilistic approach to infer the PofO that does not require prior knowledge of genealogy. For each target individual, our model (i) identifies second- and third-degree relatives and groups them into parental groups, (ii) uses Identity-By-Descent (IBD) sharing and phasing to assign the target's alleles to parental groups, and (iii) probabilistically labels parental groups as maternal or paternal. To do so, we leverage whole genome sequencing of the mitochondrial DNA to label maternal groups, and chromosome X and Y in males to label maternal and paternal groups, respectively. We use parents-offspring trios to assess the accuracy of this approach.
We use this method to probabilistically infer the PofO of 80,821 individuals in the UK Biobank, representing one of the largest datasets used to discover PofO effects so far. We scanned 99 phenotypes for PofO effects and replicated more than 90% of the loci known for PofO effects across five studies, such as the maternal effect at the MEG3/DLK1 locus on platelet phenotypes. Overall, we discovered more than 100 putative PofO effects. The strongest effect is a novel maternal effect on telomere length at the TERT locus, known to be associated with telomere length, cancer and aging. This association provides further evidence on the imprinting mechanism regulating telomere length. In addition, we found a paternal effect on BMI at the SEC16B locus, an effect that also replicates with weight, hip circumference, waist circumference and basal metabolic rate. Our approach allows clarifying the PofO nature of the Gene-Environment (GxE) interaction previously found on BMI at the SEC16B locus as well as for the additive association at the TERT locus.
Leveraging the relatedness in biobank individuals allows us to infer the PofO of alleles at an unprecedented scale and definition. Notably, by applying our method on multiple biobanks, this would enable the largest meta-analysis to discover PofO effects. We demonstrate the importance of our dataset in unraveling the underlying PofO nature of known associations, such as at the TERT or SEC16B locus. To facilitate the characterization of PofO effects, our summary statistics have been made available for the community on a user-friendly website.

S03. Featured Plenary Abstract Session I

Location: Conv Ctr/West Hall A/West Building

Session Time: Tuesday, October 25, 2022, 2:15 pm - 4:00 pm

ProgNbr 016: Characterizing 54 type-2 diabetes candidate genes using CRISPR/Cas9, *in vivo* imaging and deep learning in zebrafish larvae

Authors:

E. Mujica[1], H. Zhang[1], A. Emmanouilidou[1], E. Mazzaferro[1], C. Metzendorf[1], M. Bandaru[1], N. Cook[1], J. C. Costa[1], G. Alavioon[1], K. S. Frederiksen[2], D. Djordjevic[2], S. G. Vienberg[2], A. Larsson[3], A. Allalou[4], M. den Hoed[1]; [1]The Beijer Lab. and Dept. of Immunology, Genetics and Pathology; Uppsala Univ. and SciLifeLab, Uppsala, Sweden, [2]Novo Nordisk A/S, Måløv, Denmark, [3]Dept. of Med. Sci.; Div. of Clinical Chemistry; Uppsala Univ., Uppsala, Sweden, [4]Dept. of Information Technology; Div. of Visual Information and Interaction; Uppsala Univ. and SciLifeLab, Uppsala, Sweden

Abstract:

INTRODUCTION: Genome-wide association studies identified hundreds of loci that are associated with type-2 diabetes risk, but for most, candidate genes remain uncharacterized. Here we functionally characterize 54 candidate genes using CRISPR/Cas9, *in vivo* imaging and deep learning-based image analysis. **METHODS:** Human genes were characterized one-by-one, by jointly targeting their zebrafish orthologues using CRISPR/Cas9, in fertilized eggs from parents with transgenically expressed, fluorescently labeled β-cell nuclei *(Tg(-1.2ins:H2B-mCherry))* and hepatocytes *(Tg(fabp10a:EGFP)).* On days 9-11 post-fertilization, we acquired - in each larva - optical sections of the pancreatic islet and liver using fluorescence microscopy, as well as whole-body images in bright field. Relevant traits were quantified in imaging data using deep learning. Larvae were then sacrificed and homogenized for enzymatic assessment of glucose, LDLc, triglyceride and total cholesterol levels, as well as to distinguish crispants from controls using a fragment length analysis. **RESULTS:** By imaging 5950 CRISPR/Cas9 founders and controls for 54 genes, we identified ten genes for which crispants had an altered β-cell mass, ten genes that influenced β-cell insulin expression, and seven genes - including *pdx1* - that affected both traits. Perturbation of thirty genes influenced fat accumulation in the liver. **CONCLUSION:** Systematically characterizing candidate genes for a role in a range of type-2 diabetes-related traits using zebrafish larvae can improve our understanding of disease etiology, help prioritize genes for further in-depth characterization, and begin to shed light on the mechanisms by which these genes influence disease risk.

S03. Featured Plenary Abstract Session I

Location: Conv Ctr/West Hall A/West Building

Session Time: Tuesday, October 25, 2022, 2:15 pm - 4:00 pm

ProgNbr 017: Integrated multiomic analyses reveal LSD genes as a rich source of FXTAS modifiers

Authors:

K. Shelly, H. Seong, J. Lim, E. Allen, P. Jin; Emory Univ., Atlanta, GA

Abstract:

Fragile X-associated tremor/ataxia syndrome (FXTAS) is an adult-onset neurodegenerative disorder characterized by poorly coordinated movement, tremor, and parkinsonism. This disorder is associated with degeneration of the cerebellum and has been shown to be caused by the expansion of a CGG-repeat tract in the 5' untranslated region (UTR) of the *FMR1* gene. Typically, individuals bear alleles with 5-54 CGG repeats; however, individuals carrying a premutation allele contain 55-200 CGG repeats. FXTAS is predominantly seen in males over the age of 50 where 1:450 males harbor the genetic change in the *FMR1* gene, and approximately 40% of them develop the condition. This incomplete penetrance of FXTAS creates a barrier for investigation of the molecular factors leading to the disorder. To understand the genetic and metabolomic contribution to disease penetrance, we performed whole genome sequencing (WGS) and untargeted metabolomics on *FMR1* premutation (PM) carriers with symptom onset before age 65 and PM controls without symptoms by age 68. We identified the genes with the highest variant burdens and metabolites with differential abundance between groups and overlapped those data with the 85 known genes associated with lysosomal storage disorders (LSDs), to perform a genetic modifier screen in *Drosophila*. Consistent with increasing data from the neurodegeneration field at-large, we found that LSD genes were a rich source of genetic modifiers that modulate rCGG-mediated neurodegeneration. By crossing 142 fly RNAi lines corresponding to human LSD genes with flies expressing fragile X premutation CGG repeats, we identified 38 LSD genes that could enhance rCGG neuronal toxicity and 2 genes suppressing rCGG toxicity. These modulators of rCGG neuronal toxicity were enriched for glycosaminoglycan degrading enzymes including enhancers *ARSB*, *GALNS*, *GLB1*, *IDS*, and *IDUA* as well as the suppressors *GUSB* and *BTD*. *GALNS* was also associated with presence of disease symptoms in PM carriers in analyses of WGS data. Metabolomic profiles from the same cohort showed differential abundance of glycosaminoglycans, chief among these was the *GALNS* substrate N-acetyl-D-galactosamine, which was consistently decreased in PM carriers with FXTAS symptoms. Substrates for *ARSB*, *IDUA* and *GUSB* were also differentially abundant in symptomatic FXTAS males vs controls and analyses are ongoing. Cumulatively, we found LSD-associated genes to be a rich source of FXTAS modulators and *GALNS* specifically to be a promising therapeutic target for FXTAS.

S03. Featured Plenary Abstract Session I

Location: Conv Ctr/West Hall A/West Building

Session Time: Tuesday, October 25, 2022, 2:15 pm - 4:00 pm

ProgNbr 018: Tissue-specific dynamic eQTLs in response to a high cholesterol, high fat diet in baboons

Authors:

W. Lin[1], O. Allen[1], G. Li[2], E. Quillen[2], M. Mahaney[3], J. VandeBerg[3], A. Comuzzie[4], S. Cole[4], D. Newman[4], S. Birnbaum[4], J. Glenn[4], S. Puppala[2], K. Spradling-Reeves[2], J. Chan[2], M. Olivier[2], M. Abney[1], J. Wall[5], L. Cox[2], Y. Gilad[1]; [1]Univ. of Chicago, Chicago, IL, [2]Wake Forest Univ. Sch. of Med., Winston-Salem, NC, [3]Univ. of Texas Rio Grande Valley, Edinburg, TX, [4]Texas BioMed. Res. Inst., San Antonio, TX, [5]HIBio, San Francisco, CA

Abstract:

Despite increasing awareness that gene-by-environment (GxE) interactions explain a large portion of the missing heritability in complex diseases, experimental challenges in human studies continue to significantly limit the ability to characterize GxE effects. Model organisms, including non-human primates, offer a way to study GxE interactions in a controlled environment. Due to their remarkable genetic and physiological similarities to humans, baboons have been used to model a variety of complex human diseases, including dyslipidemia, obesity, osteoporosis and atherosclerosis.
To explore the effects of diet on metabolic traits, we obtained liver, muscle, and adipose tissue biopsy samples from 111 baboons (64M, 47F) before and after they were fed a high cholesterol, high fat (HCHF) diet for two years. Using RNA-sequencing data from 589 high-quality tissue samples, we obtained transcriptional profiles from each tissue at the two time points. We then used these data to examine differences in gene regulation following the HCHF diet, and map response cis-eQTLs following the diet. Across the three tissues, we identified 6941 differentially expressed (DE) genes (FDR = 0.05) following the HCHF diet. Gene ontology and pathway enrichment analysis indicates that these DE genes are enriched in tissue-specific functions including metabolic processes and immune responses. We proceeded by mapping cis-eQTLs, and identified 8566 genes with at least one cis-eQTL (i.e., cis-eGenes). We found that more cis-eQTLs were shared between time points (74-79%) than across tissues (42-46%). We also identified 1119 diet-response eGenes, namely genes where a nearby genotype is associated with variation in gene expression either before or after the HCHF diet, but not in both. We also characterized sex-differentiated genetic effects on gene expression and identified 103 sex-biased eGenes that show robust diet-responsiveness or tissue-specificity.
Our findings nominate candidate genes for follow-up and provide novel insights into the etiology metabolic and cardiovascular diseases. As the first eQTL mapping study in baboons, our work demonstrates that the baboon is a powerful model for understanding how genetic and environmental factors influence susceptibility to human complex diseases.

S36. Featured Plenary Abstract Session II

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 5:15 pm - 7:00 pm

ProgNbr 221: Multi-omic profiling of 204,494 cells and nuclei from human islet donors under basal and stimulatory conditions nominates causal cell types, genes, and regulatory element contexts at T1D GWAS loci

Authors:

C. Robertson[1], R. Albanus[2], B. Li[3], Y. Han[3], N. Manickman[1], N. Narisu[4], M. Erdos[4], F. Collins[4], S. Chen[3], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Washington Univ. in St. Louis, St. Louis, MO, [3]Weill Cornell, New York, NY, [4]NHGRI/NIH, Bethesda, MD

Abstract:

Progressive loss of insulin production by pancreatic beta cells, due to beta cell death or dysfunction, is the defining feature of type 1 diabetes (T1D). Genetic variants influencing T1D susceptibility alter gene regulatory responses to environmental stressors in immune and endocrine cell types. Studies examining genetic and environmental risk factors for T1D in human pancreatic islets have been limited. We exposed islets from healthy donors (n=9) to two conditions with established relevance to T1D pathogenesis: 1) proinflammatory cytokines, and 2) Coxsackie B4 virus (CVB4). We also examined islets from donors positive for islet antigen autoantibodies (AAB^+ donors, n=4), a precursor of T1D. We profiled chromatin signatures (single nuclei ATAC-seq, n=129,644 nuclei), gene expression in whole cells (scRNA-seq, n=71,375 cells) and nuclei (snRNA-seq, n=79,747), and enhancer activity (eRNAs measured with 5' snRNA-seq, n=74,850 nuclei). These included 79,747 nuclei with joint modality profiling, providing ATAC-seq and gene expression on the same nuclei. In total, four molecular profiling modalities will be integrated across 204,494 nuclei and cells. After stringent quality control, we identified ten major cell types demonstrating expression of expected marker genes. Through cross modality integration, we further confirmed expected chromatin activity in marker gene regulatory regions. The most abundant cell type was ductal cells (35%), but collectively islet endocrine cell types (alpha, beta, delta, and gamma cells) represented 34% of nuclei and cells. T1D risk variants were significantly enriched in accessible chromatin (snATAC-seq) from multiple islet cell types, including immune, endocrine, acinar, ductal, and stellate cells (fGWAS log enrichment > 1.53). Within immune cell types, T1D risk variants were more enriched in differentially accessible regions after cytokine treatment (cytokine vs control) or autoantibody positivity (AAB^+ vs healthy). Integrating fine-mapped T1D credible sets with chromatin accessibility profiles, we quantified the degree to which credible set posterior probabilities are concentrated in cell type-specific accessible chromatin. These analyses suggested endocrine-specific mechanisms for multiple T1D GWAS signals, including signals near *INS*, *DLK1*, *TOX*, *RASGRP1*, and *GLIS3*. Functional validation of regulatory mechanisms at prioritized variants in these loci is underway. We are now analyzing single nucleus islet eRNA profiles, an important component of gene regulation, which will provide additional insights about mechanisms at T1D loci.

S36. Featured Plenary Abstract Session II

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 5:15 pm - 7:00 pm

ProgNbr 222: Polygenic score performance varies across the continuum of genetic ancestry in all human populations

Authors:

Y. Ding[1], K. Hou[1], A. Pimplaskar[1], Z. Xu[1], E. Petter[1], F. Privé[2], B. Vilhjalmsson[2], L. Olde Loohuis[1], B. Pasaniuc[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Aarhus Univ., Aarhus, Denmark

Abstract:

Recent work has highlighted the limited portability of polygenic scores (PGS) across different groupings of individuals (such as genetic ancestries, socio-economic status, sex or age) contributing to inequities in PGS applications. PGS performance has typically been assessed using a single statistic (e.g., R2) across "populations" of individuals. However, population-level metrics of PGS performance ignore inter-individual variation and can therefore be problematic when assessing individual level performance.

In this work we investigate inter-individual PGS performance (error/bias) across the continuum of genetic ancestry (within and across traditionally defined continental ancestries). We show that PGS performance varies across individuals, even within traditionally defined "homogeneous" groups of individuals with similar genetic ancestry and demographic contexts. We leverage recent methods that measure PGS performance at individual level (Ding et al Nat Genet 2022) to show that PGS performance decreases individual-to-individual in all considered populations linearly with the genetic distance from the population used to train PGS models along a continuum of genetic ancestries. We introduce methods to calibrate PGS across the genetic continuum that leverages individual-level PGS uncertainty to estimate a metric of individual-level PGS accuracy (R_i^2) that can be used for PGS applications. In simulations based on real genotypes from UKBB, we show that the average individual-level PGS accuracy at each genetic distance percentile is highly associated with the correlation between PGS and genetic value.

We use data from a large and diverse EHR-linked biobank at UCLA (ATLAS, N=60K) jointly with the UK Biobank (UKBB, N=500K) to evaluate the dependency of PGS performance along the genetic continuum across a broad set of 87 traits and diseases. We observe a high correlation between genetic distance and PGS performance across all 87 traits and in all populations (r = 0.87-0.98). For example, when rank-ordering individuals according to genetic distance, a PGS trained in UKBB has 27% lower accuracy in bottom decile (vs top decile) of European-ancestry individuals; notably the bottom decile of European-ancestry individuals also showed lower accuracy than top decile of Admixed American individuals. PGS themselves are also correlated with genetic distance within each traditionally-defined continental ancestries for 82 out of 87 traits further emphasizing the impact of genetic ancestry on PGS performance. In summary, our results showcase the need for accounting for the continuum of genetic ancestry for any practical application of PGS.

S36. Featured Plenary Abstract Session II

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 5:15 pm - 7:00 pm

ProgNbr 223: Rapid and scalable preclinical evaluation of personalized antisense oligonucleotides using organoids derived from rare disease patients

Authors:

S. T. Younger[1,2,3], J. C. Means[1], D. A. Louiselle[1], B. S. Koseva[1], T. Pastinen[1,2,3]; [1]Children's Mercy Kansas City, Kansas City, MO, [2]Univ. of Missouri-Kansas City Sch. of Med., Kansas City, MO, [3]Univ. of Kansas Med. Ctr., Kansas City, KS

Abstract:

Personalized antisense oligonucleotides (ASOs) have achieved positive results in the treatment of rare genetic disease. As clinical sequencing technologies continue to advance, the ability to identify rare disease patients harboring pathogenic genetic variants that may be amenable to this therapeutic strategy is likely to improve. To support this expanded patient population, we have established a platform to facilitate the rapid characterization of preclinical ASO leads. We developed a highly-efficient and scalable pipeline for the derivation of iPSCs. Our noninvasive iPSC reprogramming method requires as few as 300K PBMCs (commonly available from prior genetic testing) and is typically complete within 4-6 weeks. A parallelized format enables the simultaneous reprogramming of dozens of patient samples, allowing the straightforward generation of >100 patient-derived iPSC lines in under 3 months with >90% reprogramming success rates. Using this pipeline we generated patient-derived iPSCs from a panel of Duchenne muscular dystrophy (DMD) patients. Included in this panel was one patient harboring a structural deletion of exons 46-53 of the *DMD* gene, a deletion amenable to treatment with the FDA-approved ASO Casimersen. Patient-derived iPSCs treated with 2'-O-methyl ASOs matching the sequence of Casimersen showed restoration of DMD protein expression within 5 days. Also included in our DMD patient panel was a pair of siblings both harboring a deep intronic variant in the *DMD* gene that gives rise to a novel splice acceptor site, incorporation of a cryptic exon, and premature transcript termination. We designed patient-specific 2'-O-methyl ASOs targeting the intronic variant and observed restoration of DMD protein expression in iPSC lines derived from both patients. To further validate the therapeutic potential of our patient-specific ASOs we differentiated patient-derived iPSCs into cardiac organoids, a 2 week process, and evaluated organoid function. Cardiac organoids from DMD patients displayed slow rates of contraction and abnormal fluctuations in calcium levels as measured by a fluorescent calcium indicator dye. Treatment with patient-specific ASOs, either prior to organoid differentiation or post-differentiation, restored both contraction rates and calcium levels to those of cardiac organoids derived from healthy iPSCs. Overall, our platform required under 8 weeks of hands-on time from sample acquisition to functional validation of patient-specific ASOs. Moreover, this platform is generalizable to any patient-derived cellular models that can be functionally profiled using fluorescence-based assays, including brain organoids.

S36. Featured Plenary Abstract Session II

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 5:15 pm - 7:00 pm

ProgNbr 224: A genome-wide atlas of recurrent repeat expansions in human cancer

Authors:

G. Erwin[1], G. Gursoy[2], R. Al-Abri[1], A. Suriyaprakash[1], E. Dolzhenko[3], K. Zhu[1], C. R. Hoerner[1], L. Ramirez[1], R. Yuen[4], A. Fan[1], J. Leppert[1], M. Eberle[3], M. Gerstein[2], M. Snyder[1]; [1]Stanford Univ., Stanford, CA, [2]Yale Univ., New Haven, CT, [3]Illumina, San Diego, CA, [4]The Hosp. for Sick Children, Toronto, ON, Canada

Abstract:

Expansion of a single repetitive DNA sequence, termed a tandem repeat (TR), is known to cause more than 50 human diseases. However, repeat expansions are often not explored beyond neurological and neurodegenerative disorders. In some cancers, mutations accumulate in short tracts of TRs (STRs), a phenomenon termed microsatellite instability (MSI); however larger repeat expansions have not been systematically analyzed in cancer. Here, we identified TR expansions in 2,622 cancer genomes, spanning 29 cancer types. In 7 cancer types, we found 160 recurrent repeat expansions (rREs); most of these (155/160) were subtype specific. We found that rREs were non-uniformly distributed in the genome with an enrichment near candidate cis-regulatory elements, suggesting a role in gene regulation. One rRE located near a regulatory element in the first intron of UGT2B7 was detected in 34% of renal cell carcinoma samples and was validated by long-read DNA sequencing. Moreover, targeting cells harboring this rRE with a rationally designed, sequence-specific DNA binder led to a dose-dependent decrease in cell proliferation. Overall, our results demonstrate that rREs are an important but unexplored source of genetic variation in human cancers, and we provide a comprehensive catalog for further study.

S36. Featured Plenary Abstract Session II

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 5:15 pm - 7:00 pm

ProgNbr 225: An oligogenic inheritance test discovers novel risk genes and interactions in congenital heart defects

Authors:

M. Pittman[1,2], K. Lee[1,3], D. Srivastava[1,3], K. S. Pollard[1,2,4]; [1]Gladstone Inst.s, San Francisco, CA, [2]Dept. of Epidemiology & Biostatistics, Univ. of California, San Francisco, CA, [3]Dept. of Pediatrics, Univ. of California, San Francisco, CA, [4]Chan Zuckerberg Biohub, San Francisco, CA

Abstract:

Congenital heart defects (CHD) are the most common birth defect, occurring in nearly 1% of live births and affecting millions of infants every year. Given its high heritability estimates of 70-90%, researchers anticipate that patient genetic data will be useful in determining genes and loci that contribute to CHD risk; however, few causal variants have been identified. Further, individuals who carry damaging mutations in known risk genes often demonstrate variable phenotypes even within the same family, indicating a probable role for incomplete penetrance and genetic modifiers within the complex etiology of CHD. Recent studies have demonstrated cases of oligogenic inheritance of CHD, such that the combination of a few high-effect damaging variants causes a phenotype, while carriers of a single or incomplete subset of these variants remain unaffected. Often these variants are found in genes whose products interact in protein complexes or key pathways in cell differentiation and cardiogenesis. To explore oligogenic causes of CHD without assessing billions of variant combinations, we developed an efficient, simulation-based method to detect gene sets that carry damaging variants in probands at a higher rate than expected given parental genotypes. We implemented this extension to the transmission disequilibrium test in a software called Gene Combinations in Oligogenic Disease (GCOD) and applied it to a cohort of 3382 CHD trios with exome sequencing. This analysis detected 353 high-confidence risk genes in 202 digenic pairs and 244 higher-order gene sets for which damaging variants appear together in multiple probands, but rarely or never appear in combination in their unaffected parents. We use a knock-down model in mice to validate the interaction of the digenic pair *GATA6* and *POR* in Persistent Truncus Arteriosus. We show that the novel oligogenic risk genes and putative modifier genes cluster in pathways related to heart development and suggest new molecular disease mechanisms, such as arylsulfatase activity and de novo nucleotide biosynthesis. Our data suggests that a sizable minority of CHD patients (10%) carry rare damaging variants in putative oligogenic sets, in combinations that were not found in unaffected family members or sequenced controls.

S53. Featured Plenary Abstract Session III

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 5:15 pm - 7:00 pm

ProgNbr 335: Systematic analysis of nonsense-mediated decay escaping protein-truncating variants in 97,728 clinical exomes identifies new Mendelian disease genes

Authors:

R. Torene[1], F. Millan[1], Z. Zhang[1], M. Stephen[1], M. Oetjens[2], S. Myers[2], D. Ledbetter[3], C. Martin[2], K. Mitchell[4], T. Brandt[1], K. Retterer[1]; [1]GeneDx, Gaithersburg, MD, [2]Geisinger, Danville, PA, [3]Unified Patient Network, Inc., Amelia Island, FL, [4]Trinity Coll., Dublin, Ireland

Abstract:

Background
Protein truncating variants (PTV) near the 3' end of genes may escape nonsense-mediated decay (NMD). PTVs in the NMD-escape region (PTVesc) have been identified as a mechanism of disease, however, no large-scale analysis has systematically evaluated genomic variants in these regions in patients with rare disease.
Methods
We performed a retrospective analysis of clinical exomes from 97,728 individuals, including 29,031 parent-child trios referred for neurodevelopmental disorders (NDD) to identify high-confidence PTVesc *de novo* mutations (DNMs).
Results
We identified 1,343 PTVesc DNMs in 841 genes. There were 97 genes with ≥2 such variants that were significantly enriched (binomial $p < 0.001$); 38 were significant after Bonferroni correction. While PTVesc enriched genes tend to have high pLI (ks-test $p < 10^{-25}$), 25/97 genes have pLI < 0.8. The 97 genes included some with dominant PTVesc previously described, e.g., *SEMA6B*, *PPM1D*, and *KAT6A/B*. Of note, we identified 32 candidate PTVesc enriched genes not previously reported to cause Mendelian disease. After pairwise semantic similarity scoring (https://github.com/GeneDx/phenopy), we identified 4 novel candidate genes with high phenotypic similarity: *DAGLA*, *RAB1A*, *HDAC2*, *MSL2*. We identified 4 patients with PTVesc DNMs in *DAGLA* and distinctive phenotypes including ataxia, developmental delay, and gaze anomalies. The variants truncate the phosphorylated tail of the DAGLα protein which is expected to dysregulate its activity. *DAGLA* has a pLI=0.9999 and no NMD-eligible PTV DNMs were found. We also identified 2 patients with PTVesc DNMs in *RAB1A,* both with intellectual disability and spasticity. One of the DNMs was found in 2 additional cases, each inherited from an affected mother. PTVesc in *RAB1A* removes the C-terminal prenylation sites necessary for interaction with VAMP1. *VAMP1* is a known spastic ataxia disease gene. Together, these data suggest that PTVesc in *DAGLA* and *RAB1A* are responsible for two novel motor syndromes. Nine cases with syndromic NDD had PTVesc DNMs in *HDAC2* (n=2) or *MSL2* (n=7). Both genes are histone modifiers, a known mechanism of NDD pathogenesis. We identified 9 more *MSL2* cases in non-trios that strengthen this finding.
Conclusions
Using a large-scale, systematic analysis of DNMs, we identified 32 candidate genes with PTVesc as a mechanism of disease defining new disease genes, 4 of which have high phenotypic similarity. The range of pLI indicates PTVesc diseases may be hypomorph, antimorph, or neomorph and functional studies are warranted. This study illustrates the importance of examining atypical mechanisms of disease to increase the diagnostic yield of genetic testing.

S53. Featured Plenary Abstract Session III

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 5:15 pm - 7:00 pm

ProgNbr 336: The correlation between CpG methylation and gene expression is driven by sequence variants

Authors:

B. Sigurpalsdottir[1,2], O. Stefansson[1], S. Rognvaldsson[1], K. Juliusson[1], G. Sveinbjornsson[1], G. Halldorsson[1], B. Gunnarsson[1], D. Beyter[1], H. Jónsson[1], S. Gudjonsson[1], T. Olafsdottir[1], S. Saevarsdottir[1], M. Magnusson[1,3], S. Lund[1], V. Tragante de O[1], A. Oddsson[1], M. Hardarson[1,2], H. Eggertsson[1], R. Gudmundsson[1], M. Frigge[1], F. Zink[1], H. Holm[1], H. Stefansson[1], T. Rafnar[1], I. Jonsdottir[1,3,4], P. Sulem[1], A. Helgason[1,3], D. Gudbjartsson[1,3], B. Halldórsson[1,2], U. Thorsteinsdottir[1,3], K. Stefansson[1,3]; [1]deCODE genetics, Reykjavík, Iceland, [2]Reykjavík Univ., Reykjavík, Iceland, [3]Univ. of Iceland, Reykjavík, Iceland, [4]Landspítali Univ. Hosp., Reykjavík, Iceland

Abstract:

Genomic regions where CpGs are unmethylated typically correspond to regulatory sequences bound by transcription factors (TFs). CpG methylation rates of these regulatory sequences are known to correlate with gene expression, but whether sequence variants that correlate with CpG methylation contribute to this relationship is currently not well understood. We determined haplotype specific methylation rates of 7.4 million CpGs using nanopore sequencing of whole blood genomes from 4,135 individuals. We identified 76,299 methylation depleted regions (MDRs) where three or more consecutive CpGs are unmethylated on at least one haplotype. Variation in haplotype specific CpG methylation rates of 24,844 MDRs (33%) associates with 24,307 cis-acting sequence variants which we refer to as allele-specific methylation QTLs (ASM-QTLs). By performing RNA sequencing of whole blood we showed that in nearly all instances where we found an association between CpG methylation and gene expression, we identified an ASM-QTL in association with methylation of those same CpGs. Furthermore, we show that it is the ASM-QTL that affects the gene expression rather than the CpG methylation. ASM-QTLs were enriched by 31.2-fold (95%CI:21.9-42.3) among sequence variants associating with variability in hematological traits which is comparable to that of missense variants, showing that ASM-QTLs are important functional units in the non-coding genome. Our findings demonstrate that most of the effects of ASM-QTLs on expression of genes is mediated through mechanisms separate from CpG methylation, and it appears that CpG methylation is influenced more by gene expression than gene expression is by CpG methylation.

S53. Featured Plenary Abstract Session III

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 5:15 pm - 7:00 pm

ProgNbr 337: Australia's National Centre for Indigenous Genomics enabling the inclusion of First Nations peoples in genomics

Authors:

H. Patel, A. Hermes, G. Mann; The Australian Natl. Univ., Canberra, Australia

Abstract:

Large-scale projects developing perpetual community resources, such as UK Biobank, 1000 Genome Project, and the Human PanGenome Reference, play an increasingly important role in biomedical research and precision medicine. The National Centre for Indigenous Genomics (NCIG), under the Indigenous governance backed by Australian federal statutory powers, establishes genomic reference resources and biospecimen collection from Indigenous Australians. NCIG is an example of a community resource built to enable broad-scale representation of ancestrally diverse populations for equitable benefits of genomics for all humans. NCIG has established deeply trusted relationships with four Indigenous communities since 2015, whereby 684 individuals have donated their genomic data and biological samples to the growing NCIG Collection. We have begun generating reference genomes leading to telomere-2-telomere assemblies for these four communities. Indigenous reference genomes reveal >2Mb of new sequences. Similarly, whole-genome sequence data from 163 individuals using Illumina short reads and ONT long reads show distinct genetic diversity missing from global resources. Of 17 million short variants, ~22% of variants are found only in Indigenous Australians and between 10 and 31% of variants are found only in a single community due to the prolonged isolation from each other and global populations. The unique genetic diversity of Indigenous Australia warrants modifications to the genomic analysis and interpretation workflows to improve research and health outcomes. NCIG continues to expand its genomic reference resources by including more communities and integrating these resources into clinical services and healthcare through national networks.

S53. Featured Plenary Abstract Session III

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 5:15 pm - 7:00 pm

ProgNbr 338: Novel therapeutic target discovery using circulating proteins in up to 42,000 UK Biobank participants through systematic Mendelian randomization and genetic colocalization

Authors:

T. Richardson[1], S. Hu[1], **L. Chen**[1], J. Liu[2], W. Gan[1], M. Traylor[1], J. M. M. Howson[1]; [1]Genetics Dept., Novo Nordisk Res. Ctr., Oxford, United Kingdom, [2]Big Data Inst., Oxford, United Kingdom

Abstract:

The circulating proteome encompasses a wealth of novel drug targets and biomarkers which hold huge therapeutic potential. In this study, we conducted the largest genome- and exome-wide association studies to date on 1,462 plasma proteins using data from up to n=42,000 participants from the UK Biobank Pharma Proteomics Project. Findings were leveraged within a bespoke Mendelian randomization (MR) framework to incorporate findings from this unparalleled sample size, which allowed us to prioritize 139 established or emerging therapeutic targets across the cardiometabolic disease spectrum. For instance, along with strong evidence supporting known drug targets, such as *LPL* with coronary artery disease (Beta=-0.10, 95% CI=-0.14 to -0.07, $P=1.4x10^{-8}$), our systematic approach provided evidence for targets such as *UMOD* with chronic kidney disease (Beta=0.23, 95% CI=0.20 to 0.26, $P=2.3x10^{-47}$). We also undertook extensive genetic colocalization to construct a putative map of shared causal variants between circulating proteins and gene expression derived from 47 tissues. In total, 507 (36%) proteins shared a causal variant with gene expression in at least one tissue at their encoding gene's locus. Directions of effect for shared variants were typically concordant on gene expression and protein levels, with whole blood derived gene expression providing the largest proportion of concordant directionality (94%) and regional brain tissues typically providing the least agreement (e.g. amygdala=70%). One possible explanation for discordant directions of effect includes the blood-brain barrier, which may be relevant for proteins such as Parkinson's disease associated target PARK7, whose lead pQTL shared the same direction of effect with PARK7 expression in 4 tissues (esophagus mucosa, heart atrial appendage, spleen and whole blood) but the opposite direction of effect in the cerebellum. Lastly, by harnessing the exomes data from UK Biobank, we were able to characterize the contribution of rare genetic variants to circulating levels of the 1,462 proteins. Gene-burden analyses identified cis-acting rare variant signals for 692 of the proteins (47%) robust to multiple testing, as well as many trans-acting associations that allowed us to construct genetically determined protein networks which may provide granular insight into the mechanisms of therapeutic targets. Our comprehensive evaluation of common and rare variants associated with plasma proteins provides an exceptional platform to galvanise drug and biomarker discovery and validation. Findings from our efforts should help pave the way for future transformative therapies to treat and prevent disease.

S53. Featured Plenary Abstract Session III

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 5:15 pm - 7:00 pm

ProgNbr 339: A deep catalog of protein coding variation from one million individuals

Authors:

S. Balasubramanian[1], S. Chen[1], K. Sun[1], X. Bai[1], S. Bao[1], J. Backman[1], M. Kapoor[1], E. Maxwell[1], T. Joseph[1], B. Boutkov[1], A. Locke[1], D. Liu[1], L. Habegger[1], J. Staples[1], D. Sharma[1], H. Kang[1], J. Emberson[2], R. Collins[2], J. Torres[2], J. Berumen[3], P. Kuri Morales[3], J. Alegre[3], R. Tapia-Conyer[3], J. Marchini[1], G. Abecasis[1], A. Shuldiner[1], J. Reid[1], W. Salerno[1], Regeneron Genetics Center, Regeneron Research Partners; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Univ. Natl. Autónoma de México, Mexico City, Mexico

Abstract:

Rare coding variation identified from exome sequencing studies continue to reveal insights into gene function and the genetic architecture of human health. Here, we report the largest and most diverse catalogue of human coding variation to date, derived from 1,010,547 uniformly processed exomes that have been sequenced at the Regeneron Genetics Center. The One Million Exome data (1ME) includes samples from 300,000 individuals of African, South Asian, East Asian and Admixed American ancestry. In total, we identify more than 35 million high quality autosomal variants, of which 12 million are observed in only one individual and 19 million are unique to our dataset (ie, have not been seen in other large-scale studies). We found 70,553 carriers of rare homozygous putative loss of function (pLOF) variants or human knockouts (KOs) in 3,366 genes, including 1,404 KOs not previously identified in other large-scale studies. 14% of inhibitory drug targets in pre-clinical studies or clinical trials have a human KO in the 1ME dataset. Deeper phenotyping of human KOs of inhibitory drug targets could help understand their efficacy and side-effect profiles. Leveraging the large sample size, we estimated heterozygous selection constraint with more precision than previous studies and identified 1,167 highly constrained genes lacking disease annotations. In addition, we calculated missense tolerance ratios and identified novel missense-constrained regions in 514 genes that have been previously described as LOF-tolerant. We also identified highly differentiated variants by measuring the extent of allelic differentiation between 1ME sub-populations using fixation index, F_{ST}. We identify association signals with variants with high F_{st} from published GWAS studies in target populations that are not seen in the reference population where the variants are rare. Finally, we interrogated the presence of known pathogenic variants in 71 ACMG autosomal genes and discovered that 2% of sequenced individuals have an actionable variant. The 1ME data will be made available via a public variant allele frequency browser, which will enable the genomic research community to enrich existing studies of rare variation in diverse populations and improve interpretation of protein-coding variants.

