ASHG 2022 Platform 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

S14. A collection of cardiovascular and metabolic disease conundrums

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 064. A large-scale multi-ethnic mitochondrial-wide association analysis for cardio-metabolic traits

Authors:

J. Zhou[1,2], A. Jensen[1,2], M. Vujkovic[3,4], D. Samuels[5], A. Hung[6,7], S. Pyarajan[8,9], T. Hulgan[6,7], P. TSAO[10], K-M. Chang[11,4], P. Reaven[2]; [1]UCLA, Los Angeles, CA, [2]Phoenix VA Hlth.Care System, Phoenix, AZ, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [5]Vanderbilt Univ. Sch. of Med., Nashville, TN, [6]Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]VA Tennessee Valley Hlth.care System, Nashville, TN, [8]Partners Hlth.Care, Lexington, MA, [9]VA Boston Hlth.care System, Boston, MA, [10]VAPAHCS/Stanford Univ., Palo Alto, CA, [11]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA

Abstract:

The role of mitochondrial (MT) DNA (mtDNA) variation in complex diseases has been underexplored, especially in Hispanic and African ancestries. Using the Million Veteran Program (MVP), we report the largest mitochondrial genome-wide association study of cardiometabolic traits across four ancestry groups: European (EUR, N=446,244), African (AA, N=114,372), Hispanic (HIS 49,289), and Asian (ASN, N=8,119). We developed an MVP mtDNA QC procedure, which included: (1) call rate >98%; (2) confirmation by intentional duplicates; and (3) comparison with overlapping whole-genome sequencing data. Among 248 mtDNA variants, 94 (38%) did not vary among the current MVP population. Of the remaining 154 mtDNA SNPs, 141 variants were rare, with minor allele frequencies (MAF) <2%. Due to the enrichment of rare variants, gene-based rare burden association analyses were performed. Cardiometabolic traits included 365 phecodes in the endocrine/metabolic, digestive, and circulatory systems, HbA1c, and BMI assessments. All reported results below were significant after a conservative Bonferroni adjustment ($p < 0.05/$(number of gene x trait pairs tested)). We identified 8 MT gene and cardiometabolic trait associations and other 28 associations when stratifying by type 2 diabetes (T2D) and sex status. Notable findings included that among HIS ancestry, the MT-ND2 gene's rare burden was associated with coronary atherosclerosis (411.4), ischemic heart disease (411), and peripheral vascular disease (443). Interestingly, when stratifying the analysis by T2D status, these associations were present among patients with T2D but not patients without T2D. In EUR ancestry, MT-ND3 was associated with vascular insufficiency of the intestine (441) among male veterans; MT-TG and MT-RNR2 were related to disorders of magnesium metabolism (275.3) and precordial pain (418.1), respectively, in female veterans. RNR2 was associated with hypoglycemia (251.1) in AA ancestry with T2D. Finally, we assessed the interaction effects of mitochondrial genes and nuclear mitochondrial genes curated from the mitoCarta 2.0 database, using a two-stage approach. Only marginally significant genes with the same condition were followed for interaction analysis. A significant interaction between UGT1A1 and MT-RNR2 on disorders of bilirubin excretion (277.4) in EUR ancestry was identified. Our work illustrates the value of mitochondrial-wide association studies for cardiometabolic diseases of all ancestries, especially HIS and AA ancestry. Future works include validations in independent cohorts, in silico, and in vitro analyses to further elucidate these findings' biological mechanisms.

S14. A collection of cardiovascular and metabolic disease conundrums

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 065. Genome-wide association study of coronary microvascular disease assessed by stress cardiac perfusion positron emission tomography

Authors:

S. Verma[1], L. Guare[1], B. Irving[1], Regeneron Genetics Center, Penn Medicine BioBank, M. Levin[2], M. Guerraty[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Background: Coronary Microvascular Disease (CMVD), defined as a disease of the coronary pre-arterioles, arterioles, and capillaries, accounts for 30-50% of the burden of ischemic heart disease (IHD). However, little is known about the pathogenesis, and there are no directed therapies for CMVD. Perfusion cardiac PET allows for the quantitative assessment of myocardial blood flow (MBF) at rest and stress and MBF reserve (MBFR) and represents the most clinical advanced non-invasive imaging modality for CMVD. Here we perform the first genome-wide association study (GWAS) for CMVD using cardiac PET MBF parameters.Methods: Rest and stress MBF and MBFR were measured using Rubidium-82 cardiac perfusion PET stress tests which were obtained as part of routine clinical care. We performed a GWAS in 383 samples of European (EUR) and 539 samples of African ancestry (AFR) in the Penn Medicine Biobank using MBF values and REGENIE software. Models were adjusted for age, sex, age^2, age x sex, 5 principal components, and batch corresponding to phenotype extraction. Additionally, fine mapping was performed for loci with P-value < 1e-05 using the implementation of SuSiE in the polyfun package by integrating gene expression and functional annotations from heart ventricular, atrium, aorta, and coronary artery tissues in the EpiMap database.Results: We identified 10 variants in AFR ancestry for MBFR at p-value < 5e-7. Among the top hits, our results identified novel signals in variants mapping to *LOC105374051, UNC13A, RPL21P130, ATP6V0A2, CCSER1, and MDFI*, among others. We also identified 4 variants in AFR and 1 in EUR for rest MBF and 3 variants in AFR and 1 in EUR for stress MBF phenotypes, including loci near *ARHGAP18 and ARHGAP22*. Fine mapping identified 4 loci in AFR for MBFR including *NRP1, ERC1, RN7SKP148, and SCL15A4*. In sex-stratified analyses, we found that variants were more strongly associated with PET-derived MBF phenotypes in AFR women than in AFR men and in EUR men relative to EUR women. Conclusions: Our study identified several variants associated with perfusion PET MBF parameters in populations of EUR and AFR ancestry, some of which represent putative CMVD loci. This work yields novel insights into the mechanisms of CMVD.

S14. A collection of cardiovascular and metabolic disease conundrums

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 066. Genome-wide analysis reveals novel mechanisms underlying atrial fibrillation and the clinical utility of a polygenic predictor for cardioembolic risk

Authors:

K. Miyazawa[1], K. Ito[1], M. Ito[2], Z. Zou[3], M. Kubota[2], S. Nomura[2], H. Matsunaga[2,1], S. Koyama[4,5,1], H. Ieki[2,1], Y. Koike[6], M. Akiyama[7,1], R. Kurosawa[1], H. Yoshida[2,1], K. Ozaki[8,1], Y. Onouchi[9,1], A. Takahashi[10,1], K. Matsuda[2], Y. Murakami[2], H. Aburatani[2], M. Kubo[1], Y. Momozawa[1], C. TERAO[1], S. Oki[3], H. Akazawa[2], Y. Kamatani[2], I. Komuro[2]; [1]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [2]The Univ. of Tokyo, Tokyo, Japan, [3]Kyoto Univ. Graduate Sch. of Med., Kyoto, Japan, [4]Broad Inst., Cambridge, MA, [5]Massachusetts Gen. Hosp., Boston, MA, [6]RIKEN Ctr. for Integrative Med. Ctr., Yokohama, Japan, [7]Kyushu Univ., Fukuoka, Japan, [8]The Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [9]Chiba Univ. Graduate Sch. of Med., Chiba, Japan, [10]Natl. Cerebral and Cardiovascular Ctr., Suita, Japan

Abstract:

Atrial fibrillation (AF) is a common cardiac arrhythmia with an estimated over 30 million people affected worldwide. Genome-wide association studies (GWASs) of AF have identified over 100 genetic loci, while the majority of variants detected by GWAS lie on non-coding regions so that their functionalities are not fully elucidated. Therefore, we performed a large-scale Japanese GWAS comprising 9,826 AF cases among 150,272 individuals, followed by a trans-ancestry meta-analysis of over 1 million individuals including 77,690 cases, and discovered 35 novel genome-wide significant loci. To explore the pleiotropic effect of AF-associated loci, we performed colocalization analysis with quantitative trait loci, which demonstrated that they were significantly colocalized with kidney-related, blood pressure-related, and metabolic-related traits. Furthermore, we performed a transcriptome-wide association study leveraging the gene expression data from GTEx, which prioritized 132 and 127 candidate causal genes in the atrial appendage and left ventricle, respectively. Particularly, we found IL6R as one of the causal genes, which was previously unreported in the genetic analysis of AF. Additionally, we assessed transcription factors enriched in AF-associated loci using the ChIP-seq Atlas and identified a novel transcription factor. The experiment of human-induced pluripotent stem cell-derived cardiomyocytes (iPSCMs) revealed the functional inhibition of the transcriptional factor showed a decrease in the gene expression of ion channels and sarcomere genes and revealed a trend toward decreased spontaneous beating rate, notable irregularity, and prolonged contraction duration. Finally, we constructed polygenic risk scores (PRS) from the trans-ancestry meta-GWAS and observed that individuals with the top 1% PRS were estimated to be approximately 4 years younger in AF onset compared to the remaining individuals. Intriguingly, we found a significant association of higher AF-PRS with increased cardioembolic stroke risk among undiagnosed AF patients, suggesting that it may play an important role in the early detection of subclinical AF. Our results provide novel biological insights into AF genetics and broaden its clinical applications.

S14. A collection of cardiovascular and metabolic disease conundrums

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 067. Pathway and cell state specific polygenic risk scores shed light into the genetic basis of coronary artery disease

Authors:

M. Kaikkonen[1], T. Örd[1], T. Lönnberg[2], H. Niskanen[1], M. Maria[1], K. Ounap[1], P. Singha[1], A. Ravindran[1], M. Pashupati[3], A. Einari[1], S. Ylä-Herttuala[1], T. Lehtimäki[3]; [1]Univ. of Eastern Finland, Kuopio, Finland, [2]Univ. of Turku, Turku, Finland, [3]Univ. of Tampere, Tampere, Finland

Abstract:

Rationale: Atherosclerotic coronary artery disease (CAD) development encompasses endothelial dysfunction, vascular infiltration of immune cells and smooth muscle cell dedifferentiation. However, the regulatory mechanisms underlying the disease associated changes in gene expression and their relevance in mediating the genetic risk for CAD is largely unknown.
Objective: Here, we investigated the similarities and differences in gene signatures and epigenetic profiles of disease associated cell states among all major lesional cell types. We further interrogate the contribution of cell state markers and pathways to the genetic risk of coronary artery disease.
Method and Results: We performed single-cell RNA and ATAC sequencing to generate transcriptomic and epigenomic maps of mouse and human atherosclerotic lesion. We identified thirteen disease-associated cell states, each representing distinct gene signatures. Functional gene annotation, ligand-receptor prediction and transcription factor inference was used to identify common pathways defining disease states. This information was used to derive pathway and cell state specific polygenic risk scores. Our results demonstrated that cholesterol transport, lipid localization, blood vessel development and angiogenesis gene sets were among the top in predictive performance whereas smooth muscle cell state genes contributed most to SNP-based heritability of CAD. In line with this, variants within smooth muscle cell specific regulatory elements outperform other cell states and types in PRS performance. Finally, we performed experimental fine-mapping of the CAD GWAS variants using expression quantitative trait loci analysis, massively parallel reporter assay and CRISPR-based perturbations to identify functional disease mechanisms.
Conclusions: We identified 13 expression programs that are recurrently heterogeneous within atherosclerotic lesions. We demonstrate that the gene signatures specific for atherosclerosis associated cell states contribute to SNP-based heritability of CAD and exhibit unequal polygenic contribution to genetic risk. Our analysis also provides novel insights into the functional mechanisms through which common genetic variation affects the risk of CAD.

S14. A collection of cardiovascular and metabolic disease conundrums

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 068. Discovery of novel pharmacogenomic biomarkers of clopidogrel response in African Americans using multi omics

Authors:

G. Yang[1], T. De[2], C. Alarcon[1], J. Avitia[1], P. Friedman[1], E. Nutescu[3], T. O'Brien[4], A. Barbour[4], N. Lee[4], A. Horvath[4], M. Tuck[5], D. Meltzer[6], M. Perera[1]; [1]Dept. of Pharmacology, Feinberg Sch. of Med., Northwestern Univ., Chicago, IL, [2]Regeneron Genetics Ctr., Chicago, IL, [3]Dept. of Pharmacy Practice, Coll. of Pharmacy, Univ. of Illinois, Chicago, IL, [4]Dept. of Pharmacology and Physiology, George Washington Univ., Washington, DC, [5]Section of Hosp. Med., DC Veteran Admin. Ctr., Washington, DC, [6]Section of Hosp. Med., Dept. of Med., Univ. of Chicago, Chicago, IL

Abstract:

Clopidogrel is a platelet P2Y12 receptor antagonist used with aspirin to avoid thrombotic events in patients with acute coronary syndrome (ACS) or those undergoing percutaneous coronary intervention (PCI). Non-response to clopidogrel results in recurring ischemic episodes while on medication, with African Americans suffering disproportionately. The aim of this study was to identify pharmacogenomic biomarkers of clopidogrel resistance in African American patients, who have been left out of pharmacogenomics studies thus far.
Samples were obtained through African American Cardiovascular Pharmacogenomics Consortium (ACCOuNT). All patients were treated with 75mg/day of clopidogrel. P2Y12 reaction units (PRU) values were obtained after 15 days of treatment. High on-treatment platelet reactivity (HTPR) was defined as patients with PRU > 230. A total of 38 cases and 103 controls were included. A genome-wide association study was performed to identify variants associated with HTPR adjusting for age, was and principal components. RNAseq analysis of whole blood was used to identify gene expression differences between cases and controls. Functional follow-up of significant genes was carried out using shRNA gene knockdown in MEG-01 cells (megakaryocyte cell line) followed by RT-qPCR and western blotting. Additional *in silico* analyses were conducted on significantly associated genes and SNPs to determine functional loci and biological pathways.
rs7807369 on chromosome 7 was associated with HTPR ($p=5.57x10^{-9}$, OR=9.11, 95% confidence interval=3.66-22.67). This SNP is located at the intronic region of *THSD7A.* Two other SNPs in high LD with rs7807369 were located at the enhancer element region. The differential gene expression analysis using an independent clopidogrel GEO gene expression dataset validated that higher expression of *THSD7A* gene was associated with the HTPR. Differential gene expression analysis of the ACCOuNT cohort identified lower expression of *LAIR1* associated with HTPR. The knock-down of *LAIR1* gene increased *SYK* and decreased *SRC* mRNA expression, with similar results at the protein level.
Our data suggest that both *LAIR1* and *THSD7A* may be important factors in clopidogrel response in African Americans.The *LAIR1* deficient cells potentially induced the activation of PI3K/Akt pathway by activating SYK. This pathway was also reported to increase platelet aggregation and platelet count in a mouse model. *THSD7A* has been associated with venous thrombosis and coronary artery disease. Both genes are novel associations to clopidogrel response and underscore unique African ancestry variants contributing to this important clinical phenotype.

S14. A collection of cardiovascular and metabolic disease conundrums

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 069. Identification of *BRAF* as a novel gene for autosomal dominant dilated cardiomyopathy

Authors:

C. Lam[1], E-y. Law[2], S. Chang[2], C. Yung[3], K-y. Fan[3]; [1]The Univ. of Hong Kong, Pokfulam, China, [2]Queen Mary Hosp., Hosp. Authority, Hong Kong, Hong Kong SAR, China, [3]Dept. of Cardiac Med. Unit, Grantham Hosp., Hong Kong SAR, China

Abstract:

Introduction: We report a novel loss of phosphorylation missense variant p.Ser447Asn in *BRAF* and a proposed patho-mechanism for autosomal dominant dilated cardiomyopathy (DCM) associated with the *BRAF* paradox. The patient was a 34 years old man who presented with DCM and atrial fibrillation at the age of 28 years old. He also has a positive family history of sudden cardiac death (father and younger brother) of unknown etiology and the father-to-son transmission is consistent with the autosomal dominant inheritance.
Methods: Clinical whole exome sequencing (CWES) was performed using an Illumina NextSeq2000 sequencer. Read alignment (using GRCh37/hg19), variant calling (VCF), annotation, etc. were performed using a combination of commercial and in-house bioinformatics pipelines. A virtual gene panel on familial cardiomyopathy was applied on the exome and the filtered variants were interpreted by two chemical pathologists with qualifications in genetics and genomics, according to the patient's laboratory and clinical phenotype.
Results: No known pathogenic or likely pathogenic variants were found in the familial cardiomyopathy panel. Interestingly, a heterozygous variant in *BRAF*, NM_004333.4: c.1340G>A; NP_004324.2:p.Ser447Asn was identified. The p.Ser447Asn in *BRAF* was not found in population and mutation databases, i.e. gnomAD, 1000Genomes, ClinVar and HGMD. The SSDD sequence from residues 446 - 449 was critical for *BRAF* functioning and a loss of serine phosphorylation at p.Ser447 will result in a drastic reduction of basal and Ras-GTP-stimulated kinase activities. In this regard, p.Ser447Asn is defined as a class III type of *BRAF* mutation, i.e. impaired kinase activity or kinase-dead. BRAF Ser447 phosphorylation was demonstrated in both G1- and M- phases by mass spectrometry. In addition, p.Ser447Asn could lead to impairment of RAF1-BRAF heterodimer formation. In zebrafish studies, autosomal dominant DCM-associated *RAF1* mutants resulted in a hyperactivated AKT pathway in a *BRAF*-dependent manner, with the DCM phenotype being rescued by rapamycin. Taken together, the impaired formation of RAF1-BRAF heterodimer could contribute to the new gene-phenotype association of *BRAF*, i.e. autosomal dominant DCM.
Conclusion: We report a loss of phosphorylation of *BRAF* could lead to autosomal dominant DCM. The novel mechanism of DCM may open new insights for the understanding of the *BRAF* paradox and the adverse cardiotoxic effects of *BRAF* inhibitors in cancer patients, and also the potential therapeutic effect of rapamycin on DCM.

S15. Applying Mendelian randomization to complex traits

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 072. Integration of single-nucleus RNA-sequencing into Mendelian randomization reveals an adipose cell-type-origin effect of abdominal obesity on non-alcoholic fatty liver disease

Authors:

S. Lee[1], Z. Miao[1,2], M. Alvarez[1], N. Darci-Maher[1], J. N. Benhammou[3], D. Z. Pan[1,2], U. Arasu[4], T. Örd[4], D. Kaminska[5], S. Heinonen[6], M. Wabitsch[7], M. Laakso[8], V. G. Agopian[9], J. R. Pisegna[10], K. H. Pietiläinen[6,11], M. U. Kaikkonen[4], J. S. Sinsheimer[1,12], J. Pihlajamäki[5,13], P. Pajukanta[1,2,14]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at Univ. of California, Los Angeles, Los Angeles, CA, [2]Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, [3]Vatche and Tamar Manoukian Div. of Digestive Diseases and Gastroenterology, Hepatology and Parenteral Nutrition, David Geffen Sch. of Med. at Univ. of California, Los Angeles and VA Greater Los Angeles HCS, Los Angeles, CA, [4]A. I. Virtanen Inst. for Molecular Sci., Univ. of Eastern Finland, Kuopio, Finland, [5]Inst. of Publ. Hlth.and Clinical Nutrition, Univ. of Eastern Finland, Kuopio, Finland, [6]Obesity Res. Unit, Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [7]Div. of Pediatric Endocrinology and Diabetes, Dept. of Pediatrics and Adolescent Med., Univ. of Ulm, Ulm, Germany, [8]Inst. of Clinical Med., Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland, [9]Dept. of Surgery, David Geffen Sch. of Med. at Univ. of California, Los Angeles, Los Angeles, CA, [10]Dept. of Med. and Human Genetics, Div. of Gastroenterology, Hepatology and Parenteral Nutrition, David Geffen Sch. of Med. at Univ. of California, Los Angeles and VA Greater Los Angeles HCS, Los Angeles, CA, [11]Obesity Ctr., Endocrinology, Abdominal Ctr., Helsinki Univ. Central Hosp. and Univ. of Helsinki, Helsinki, Finland, [12]Dept. of Computational Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [13]Dept. of Med., Endocrinology and Clinical Nutrition, Kuopio Univ. Hosp., Kuopio, Finland, [14]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract:

Non-alcoholic fatty liver disease (NAFLD) is a chronic liver disease with heterogeneous etiology. Individuals with metabolically unhealthy obesity, characterized by accumulation of abdominal fat, are known to have higher risks for fatty liver. However, the directional biological mechanisms and cross-tissue effects of abdominal obesity and NAFLD are difficult to establish due to horizontal pleiotropy in Mendelian Randomization (MR) analyses due to confounders associated with obesity genetic signals. Here we developed a new tissue-of-origin, cell-type aware MR approach that improves interpretation of biological cross-tissue effects on NAFLD by integrating single-nucleus RNA-sequencing (snRNA-seq) data into MR through selection of biologically well-defined, cell-type-specific instrumental variables (IVs). To identify these IVs, we conducted *cis*-expression quantitative trait loci (*cis*-eQTL) analysis in 262 independent subcutaneous adipose tissue samples and overlapped the results with variants associated with waist-hip ratio adjusted for BMI (WHRadjBMI), a well-established proxy of abdominal obesity. We then performed adipose snRNA-seq (n=8) and identified 1,643 unique cell-type marker genes across 17 adipose cell-types and subcell-types. Integrating these adipose snRNA-seq data with the WHRadjBMI GWAS *cis*-eQTL results, we found 3,929 GWAS *cis*-eQTL variants regulating expression of 66 adipose cell-type marker genes. We further conducted colocalization analysis on these variants and discovered 12 non-redundant (R2<0.2) colocalized variants that target 12 genes, including *NET1*, *PTH1R*, and *ZZEF1*, which are known obesogenic genes in mouse. None of these 12 genes have previously been reported as underlying GWAS genes for abdominal obesity in WHRadjBMI GWAS, suggesting that these are novel candidate genes for abdominal obesity. Next, we incorporated the 12 colocalized WHRadjBMI GWAS *cis*-eQTL variants as IVs into MR and discovered a significant causal effect of abdominal obesity on NAFLD. Multiple MR methods provided consistent results without any evidence of pleiotropy or heterogeneity. Our snRNA-seq data and functional experiments further revealed that the majority of these 12 target genes are highly expressed in human adipocytes and differentially expressed during adipogenesis, emphasizing their tissue- and cell-type-origin effects in MR. Overall, we discovered tissue-of-origin, cell-type aware causal effects of abdominal obesity on NAFLD by integrating snRNA-seq data into MR. Our approach is generalizable for other complex disorders to improve the statistical accuracy and biological discovery potential of MR.

S15. Applying Mendelian randomization to complex traits

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 073. Mendelian randomization and colocalization characterize the impact of the plasma proteome on human complex diseases

Authors:

J. Robinson[1], **D. Baird**[2], B. Sun[2], Y. Huang[2], Biogen Biobank Team, H. Runz[2], E. Tsai[2], T. Gaunt[1], C. Whelan[2], C-Y. Chen[2]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]Biogen, Cambridge, MA

Abstract:

Plasma proteins play a key role in biological processes and often point to mechanisms underlying human complex diseases. Genetic variants associated with the abundance of plasma protein (protein quantitative trait loci; pQTLs) provide important evidence to support the functional impact of disease-associated genes. However, previous pQTL databases are limited in sample size and translation of pQTL evidence to support functional impact of specific disease-associated genes are often subject to methodological constraints. To systematically evaluate the impact of plasma proteome on human complex diseases, we employed phenome-wide Mendelian randomization (MR) and colocalization (coloc) on the human plasma proteome using pQTLs available from UK biobank (N sample=57,000) and disease outcome genome-wide association studies (GWAS) in FinnGen. Using a combination of both MR and coloc, we systematically analyzed 1,472 plasma proteins on genetic liability to diseases and traits within FinnGen. Independent cis-acting and trans-acting pQTL instruments were derived from UK Biobank European samples for the MR and coloc analysis. To address previous methodological constraints, we developed a high-throughput pipeline which aggregates MR and coloc results with relevant sensitivity analyses and functional evidence from the literature. In our cis-acting MR and coloc analysis, we tested pQTLs for 1,472 proteins across 2,272 health-related outcomes derived in FinnGen, which resulted in assessing nearly 3.4 million disease-gene pairs. As a confirmatory example, we observed an inverse association between *PCSK9* pQTLs and genetic liability to familial hypercholesterolaemia in FinnGen (OR = 0.59, 95% Confidence Interval: 0.52, 0.69; $P = 6.0 \times 10^{-12}$). We replicated this result using genetic data of low-density lipoprotein cholesterol available from the Global Lipids Genetics Consortium ($\beta = -0.64$, $P = 6.5 \times 10^{-41}$), recapitulating the effect of *PCSK9* inhibitors. Employing a novel analytic pipeline, our large-scale MR and coloc study identified numerous proteins with potential causal roles in a wide spectrum of complex diseases, providing insights to facilitate novel drug target discovery.

S15. Applying Mendelian randomization to complex traits

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 074. Proteome-wide Mendelian randomization implicates nephronectin as an actionable mediator of the effect of obesity on COVID-19 severity

Authors:

S. Yoshiji[1], G. Butler-Laporte[1], T. Lu[1], J. D. S. Willett[1], C-Y. Su[1], T. Nakanishi[1], D. Morrison[1], Y. Chen[1], K. Liang[1], M. Hultström[2], Z. Afrasiabi[3], S. Lan[3], N. Duggan[3], C. DeLuca[3], M. Vaezi[3], C. Tselios[3], X. Xue[3], M. Bouab[3], F. Shi[3], L. Laurent[3], H. M. Münter[1], M. Afilalo[1], J. Afilalo[1], V. Mooser[1], N. J. Timpson[4], H. Zeberg[5], V. Forgetta[3], S. Zhou[1], Y. Farjoun[3], J. B. Richards[1]; [1]McGill Univ., Montréal, QC, Canada, [2]Upsala Univ., Upsala, Sweden, [3]Lady Davis Inst., Montréal, QC, Canada, [4]Univ. of Bristol, Bristol, United Kingdom, [5]Karolinska Inst., Stockholm, Sweden

Abstract:

Obesity is a major risk factor for COVID-19 severity; however, its underlying mechanism is not fully understood. Considering that obesity influences the human plasma proteome, we sought to identify circulating proteins mediating the effects of obesity on COVID-19 severity by integrating proteome-wide Mendelian randomization (MR), mediation analysis, colocalization, and single-cell RNA-sequencing (scRNA-seq) analysis.
The study was conducted in the following manner:
1) Step 1 MR: We performed a two-sample MR to estimate the effect of body mass index (BMI) on circulating protein levels. For this, we used BMI genome-wide association study (GWAS) from GIANT and UK Biobank (n = 694,649) as the exposure and the 4,907 protein level GWAS from deCODE (n = 35,559) as the outcome. We found that 1,216 proteins were influenced by BMI ($P < 1.0 \times 10^{-5}$; Bonferroni correction) with no detectable horizontal pleiotropy or reverse causation.
2) Step 2 MR: Next, we evaluated the causal effects of the above-identified proteins (BMI-driven proteins) on COVID-19 outcomes, again using MR. We used cis-pQTLs of BMI-driven proteins from deCODE and COVID-19 outcomes (cases: ~122,616, controls: ~2,475,240) from COVID-19 HGI. MR revealed that a one SD increase in nephronectin (NPNT) levels was associated with severe COVID-19 (OR = 1.71, 95% CI: 1.45-2.02, $P = 1.63 \times 10^{-10}$) and COVID-19 hospitalization (OR = 1.36, 95% CI: 1.22-1.53, $P = 4.52 \times 10^{-8}$).
3) Mediation analysis: We performed an MR mediation analysis to quantify the extent to which the total effect of BMI on severe COVID-19 was mediated by plasma NPNT levels. We found that plasma NPNT levels mediated 3.7% (95 % CI: 0.7-6.7%, $P = 1.75 \times 10^{-2}$) of the effect of BMI on severe COVID-19.
4) Follow-up analysis for NPNT: To better understand whether the total levels of the NPNT protein or a specific NPNT splice isoform influenced COVID-19 severity, we performed colocalization of cis-pQTL with eQTL and sQTL GWASs from GTEx. We found that pQTL and sQTL in the lung colocalized with a posterior probability (PP) = 100.0%. However, this was not true for eQTL (PP = 7.4%), indicating that an NPNT splice isoform drove this effect. Furthermore, scRNA-seq data in lung samples of patients who died of COVID-19 showed that NPNT is expressed in fibroblast and alveolar cells ($P < 0.001$), implicating its role in fibrosis and air exchange. Finally, multivariable MR revealed that decreasing body fat mass (beta = 0.23, 95% CI: 0.13-0.33, $P = 6.86 \times 10^{-6}$) and increasing fat-free mass (beta = -0.15, 95% CI: -0.26, -0.04, $P = 5.83 \times 10^{-3}$) can lower NPNT levels and thus may improve COVID-19 outcomes.
These findings provide actionable insights into how obesity influences COVID-19 severity.

S15. Applying Mendelian randomization to complex traits

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 075. Pleiotropy testing identifies novel loci associated with lipid traits in multiple ethnicities

Authors:

N. Lorincz-Comi, G. Li, Y. Yang, X. Zhu; Case Western Reserve Univ., Cleveland, OH

Abstract:

Introduction Identifying genes that affect multiple traits, termed pleiotropy, can help us better understand the genetic etiology of complex phenotypes. The purpose of this study is to identify novel genomic loci - i.e., undetected in standard GWAS - with pleiotropy evidence for four blood lipid traits (HDL, LDL, triglycerides, and total cholesterol) in six ancestral cohorts (EUR, AFR, E. and S. Asian, HIS, trans-ancestry). **Methods** In each ancestral cohort, we used Mendelian Randomization (MR) to estimate bi-directional pairwise causal relationships between the four lipids and subsequent causal estimates to search the genome for genes with pleiotropy evidence using GWAS summary statistics for ~1.6 million total individuals (Graham et al., 2021). We additionally performed genome-wide meta-analyses between the four lipid traits and within each ancestral group. Genomic loci (1Mb windows) that were detected in pleiotropy testing or meta-analysis but not in the original ancestry-specific or trans-ancestry GWAS were considered novel. **Results** Pleiotropy testing identified 321 distinct novel loci, 134 of which were found in European, 47 in African, 35 in E. or S. Asian, 10 in Hispanic, and 129 in trans-ethnic cohorts (34 novel loci found by multiple ancestries). Between-lipid meta-analysis identified 322 novel loci, 77 of which (24%) were also found by pleiotropy testing. In total, we discovered 566 distinct novel loci. Of the 366 protein-coding genes identified by pleiotropy testing and meta-analysis, many have been previously associated with traits correlated with blood lipids. For example, in pleiotropy testing, we detected the *AFF3* gene which is associated with body mass index and Type I diabetes.
Heritability estimates ranged from 5-15%, with some differences between ancestral groups (e.g. for triglycerides, h^2=5.5% in Africans and 15% in E. Asians; P=0.004). Genetic correlations between ancestry groups and within lipids were generally strong (e.g., 0.27-0.95 for LDL). We detected bi-directional causal relationships between all pairs of lipid traits. MR causal estimates and genetic correlations were strongly correlated (r=0.91), with deviations from equality potentially due to pleiotropy. **Conclusion** Pleiotropy testing identified hundreds of novel genomic loci undetected by standard GWAS testing or meta-analysis despite their excellent power. Novel loci identified by pleiotropy testing have epidemiological plausibility based on historical evidence and may be associated with lipid-related morbidity.

S15. Applying Mendelian randomization to complex traits

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 076. Genetically proxied PCSK9 inhibition provides indications of lower prostate cancer risk: A Mendelian randomization study

Authors:

S. Fang[1], J. Yarmolinsky[1], D. Gill[2,3,4], C. J. Bull[1,5], C. M. Perks[5], the PRACTICAL consortium, G. Davey Smith[1], T. R. Gaunt[1], T. G. Richardson[1,2]; [1]MRC Integrative Epidemiology Unit (IEU), Population Hlth.Sci., Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, [2]Novo Nordisk Res. Ctr., Oxford, United Kingdom, [3]Clinical Pharmacology and Therapeutics Dept., Inst. for Infection and Immunity, St George's, Univ. of London, London, United Kingdom, [4]Dept. of Epidemiology and Biostatistics, Sch. of Publ. Hlth., Imperial Coll. London, London, United Kingdom, [5]IGF & Metabolic Endocrinology Group, Translational Hlth.Sci., Bristol Med. Sch., Learning & Res. Building, Southmead Hosp., Bristol, United Kingdom

Abstract:

Introduction: Prostate cancer (PrCa) is the second most prevalent malignancy in men worldwide. Observational studies have linked the use of low-density lipoprotein cholesterol (LDL-c) lowering therapies with a reduced risk of PrCa, which may have been potentially biased by confounding factors. Here, we evaluated the effects of genetically proxied LDL-c lowering drug targets on risk of PrCa using Mendelian randomization (MR). **Methods:** Genetic variants leveraged as proxies for the therapeutic inhibition of targets HMGCR, NPC1L1 and PCSK9 were identified based on a genome-wide association study (GWAS) of LDL-c ($P<5\times10^{-8}$) conducted by the Global Lipids Genetics Consortium (N=173,082). Effect estimates on PrCa risk were obtained from the PRACTICAL consortium (79,148 cases, 61,106 controls). Replication was performed using genetic instruments from an LDL-c GWAS conducted on males from the UK Biobank of European descent (N=201,678). Findings were additionally evaluated by colocalization and MR using liver-derived expression quantitative traits loci (eQTLs) and plasma protein QTLs (pQTLs). Additionally, we examined the role of potential mediators, including body mass index (BMI), lipoprotein A (Lp(a)) and testosterone. **Results:** MR using the inverse-variance weighted approach provided strong evidence supporting an effect of genetically mimicked PCSK9 inhibitors on lower PrCa risk (odds ratio (OR)=0.84 per drug effect equivalent to a standard deviation (SD) reduction in LDL-c, 95% confidence interval (CI)=0.74 to 0.96, $P=7.86\times10^{-3}$), whereas there was little evidence of an effect for genetically proxied HMGCR (OR=0.83, 95% CI=0.67 to 1.03, P=0.093) or NPC1L1 (OR=1.27, 95% CI=0.87 to 1.87, P=0.218) inhibition on PrCa risk. Analyses using male-stratified instruments provided consistent results. Colocalization identified strong evidence (CLPP=0.103) of a shared genetic variant (rs553741) between liver-derived *PCSK9* expression and PrCa risk. Moreover, pQTL analyses supported an inverse association between plasma PCSK9 levels and PrCa risk (OR=0.93 per SD reduction in PCSK9, 95% CI=0.87 to 0.997, P=0.04). Genetically proxied PCSK9 inhibition was strongly associated with Lp(a) levels (Beta=-0.07 SD change per drug effect equivalent to an SD reduction in LDL-c, 95% CI=-0.10 to -0.03, $P=1.44\times10^{-4}$), indicating that it may play a mediatory role along the pathway towards PrCa risk. **Conclusions:** Our study demonstrates that genetically proxied inhibition of PCSK9 is strongly associated with a lower risk of PrCa. Further evidence from clinical studies is needed to confirm this finding as well as the putative mediatory role of Lp(a).

S15. Applying Mendelian randomization to complex traits

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 077. Mendelian randomization (MR) identifies possible causal gene expression in multiple sclerosis (MS)

Authors:

J. Harley[1,2], L. Chepelev[1,2]; [1]US Dept. of VA, Cincinnati, OH, [2]Cincinnati Ed. and Res. for Veterans Fndn., Cincinnati, OH

Abstract:

Under overly optimistic assumptions, the traditional Mendelian Randomization (MR) approaches provide inference from the effect of risk variants upon gene expression to nominate causal gene regulatory mechanisms. Recent advances in MR methods control for correlated and uncorrelated pleiotropic effects, assess unmeasured confounding, and allow summary statistics in analyses. Cumulatively, these approaches tend toward relaxing the stringent assumptions made with earlier MR methods, especially when applied to sparse data. We have applied this new capacity to evaluate the possible causal impact of risk locus variants upon associated gene expression targets in Multiple Sclerosis (MS) to estimate parameters of a generalized MR model. This approach separates those gene expression effects that are associated with genetic loci and not causal from those that potentially are causal. We applied these methods to MS using the GWAS summary statistics provided by the international MS Genetics Consortium (1) and the eQTL data from Vosa et al (2). The single variant MR analysis without considering pleiotropy or confounding preliminarily identified the expression of 245 genes as potentially being causal for MS. The expression of most of these genes was eliminated as being causal when pleiotropy and confounding were incorporated into the MR analysis. Our analyses, when incorporating pleiotropy and confounding, show that the loci labeled as *ODF3B*, *TBX6*, *OS9*, *PRM1-RMI2*, *PITPNM2-MPHOSPH9*, and *NPEPPS-KPNB1* remain strong candidates for causation through expression of *TYMP*, *GDPD3*, *TSPAN31*, (*DEXI, RMI2*), *ARL6IP4*, and *TBKBP1*, respectively (Table). The complexity of HLA locus hinders confident conclusions from the analysis. This newer MR approach identifies a high proportion of false positive causal gene expression findings, as distinguished from previous approaches. The new MR approaches not only sets priorities for experimental validation, but also helps identify potentially authentic participants in genetic mechanisms of disease.
1. Int MS Genetics Consortium. Science 365:eaav7188. 2. Vosa U et al. Nature Genetics 53:1300, 2021.

S16. Considering ancestry in study populations

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 080. Genetic and phenotypic associations of whole blood gene expression in the admixed population of Greenland

Authors:

K. Hanghoej[1], L. Zilong[1], F. Stæger[1], M. Andersen[1], S. Ramne[1], S. Stinson[1], C. Santander[1], M. Jørgensen[2,3,4], C. Larsen[3], E. Zeggini[5,6], N. Grarup[1], I. Moltke[1], T. Hansen[1], A. Albrechtsen[1]; [1]Univ. of Copenhagen, Copenhagen, Denmark, [2]Copenhagen Univ. Hosp. - Steno Diabetes Ctr., Copenhagen, Denmark, [3]Ctr. for Publ. Hlth.Res. in Greenland, Odense, Denmark, [4]Steno Diabetes Ctr. Greenland, Nuuk, Greenland, [5]Helmholtz Zentrum Munich, Neuherberg / Munich, Germany, [6]Technical Univ. of Munich, Munich, Germany

Abstract:

The vast majority of GWAS and eQTL studies are based on single ancestry cohorts with European ancestry being the most common. However, studies with individuals of mixed ancestries can further improve our understanding of the genetic architecture and identification of causal variants. In this study, we analyzed 497 individuals from the Greenlandic population with genetic data and whole blood transcriptomes. Most of these individuals are admixed with an average of 25% of their genetic ancestry being European and 75% being Inuit. Based on unsupervised clustering of whole genome genetic data for these individuals, we can stratify them into eight genetically distinct groups. Interestingly, we can also map their transcriptome variance into these localities and the proportion of European ancestry. However, we show that a much larger amount of the transcriptomic variance is explained by correlation between close relatives such as siblings and to a lesser degree parent-offspring. After stringent bonferroni correction, we identifed 668749 cis-eQTL and 3404 trans eQTL. The trans-eQTL can be stratified into 25 independent groups. Of the novel nine novel trans eQTLs group most had either a cis-eQTL or missense variants in a gene coding for a transcription factor. Interestingly, we find eQTLs in close proximity to genes explain more of the gene expression variance in the mixed ancestry than either of the individual European and Inuit ancestries. In particular, we identify strong eQTL associations for variants that are almost fixed for the ancestral and derived allele in each ancestry, respectively, that would therefore require extremely large sample sizes to identify in a single ancestry cohort. As an example, we find a cis-eqtl variant to the FADS2 gene, a fatty acid desaturase, where 70% of the expression variance is explained by the variant despite being almost fixed for different alleles in each ancestry. Lastly, we observe differences in effect sizes, depending on the local ancestry and found that the Inuit has a low number of cis eQTLs than Europeans but that the ones they have individually explain more of the expression variance.

S16. Considering ancestry in study populations

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 081. Multi-ancestry genome-wide association study in >2.5 million individuals reveals distinct biological pathways driving type 2 diabetes susceptibility with heterogeneous effects across diverse population groups

Authors:

K. Hatzikotoulas[1], L. Southam[1], K. Suzuki[2], X. Yin[3], C. N. Spracklen[4], N. W. Rayner[1], R. Mandla[5,6], A. Huerta[5,6], K. M. Lorenz[7,8], H. Taylor[9,10], DIAMANTE Consortium, VA Million Veteran Program, J. M. Mercader[5,6,11], J. B. Meigs[5,11], M. Boehnke[3], M. I. McCarthy[12], M. Vujkovic[7,13], A. Mahajan[12], J. I. Rotter[14], B. F. Voight[7,8,15], E. Zeggini[1,16], A. P. Morris[1,2]; [1]Inst. of Translational Genomics, Helmholtz Zentrum München, German Res. Ctr. for Environmental Hlth., 85764 Neuherberg, Germany, [2]Ctr. for Genetics and Genomics Versus Arthritis, Ctr. for Musculoskeletal Res., Div. of Musculoskeletal and Dermatological Sci., Univ. of Manchester, Manchester, United Kingdom, [3]Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI, [4]Dept. of Epidemiology and Biostatistics, Univ. of Massachusetts-Amherst, Amherst, MA, [5]Programs in Metabolism and Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Diabetes Unit and Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [7]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [8]Dept. of Systems Pharmacology and Translational Therapeutics and Dept. of Genetics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [9]Ctr. for Precision Hlth.Res., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [10]British Heart Fndn. Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, [11]Dept. of Med., Harvard Med. Sch., Boston, MA, [12]Oxford Ctr. for Diabetes, Endocrinology and Metabolism, Radcliffe Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [13]Dept. of Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [14]The Inst. for Translational Genomics and Population Sci., Dept. of Pediatrics, the Lundquist Inst. for BioMed. Innovation (formerly Los Angeles BioMed. Res. Inst.) at Harbor–UCLA Med. Ctr., Torrance, CA, [15]Inst. of Translational Med. and Therapeutics, Univ. of Pennsylvania, Philadelphia, PA, [16]TUM Sch. of Med., Technical Univ. of Munich and Klinikum Rechts der Isar, 81675 Munich, Germany

Abstract:

To gain insight into biological processes underlying type 2 diabetes (T2D) pathophysiology across diverse populations, we aggregated genome-wide association study (GWAS) summary statistics via meta-regression (MR-MEGA) in 2,535,601 individuals (428,452 T2D cases) from five major ancestry groups: European (60.3%); East Asian (19.8%); African (10.7%); Hispanic (5.9%); and South Asian (3.3%). We identified 1,289 independent association signals at genome-wide significance ($P<5x10^{-8}$), mapping to 611 loci, of which 145 have not been previously reported. We performed a look-up of associations of T2D-risk alleles at index SNVs for each signal in published cardiometabolic trait GWAS and conducted unsupervised K-means clustering. We identified robust clusters of signals related to insulin production/processing in the pancreatic beta-cell and to mechanisms of insulin response. The insulin production/processing cluster was characterised by T2D-risk alleles associated with increased fasting glucose and HbA1c, and decreased fasting insulin, visceral adipose tissue (VAT), and abdominal subcutaneous adipose tissue (ASAT), and included associations mapping to loci with established effects on beta-cell function (*MTNR1B*, *TCF7L2*, and *SLC30A8*). The insulin response cluster could be partitioned into subsets of signals related to: obesity (increased body mass index, waist-hip ratio, body fat, and VAT/ASAT, including signals mapping to *FTO* and *MC4R*); lipodystrophy-mediated insulin resistance (increased fasting insulin and triglycerides, and decreased high-density lipoprotein cholesterol and body fat, including signals mapping to *IRS1* and *ANKRD55*); and liver lipid metabolism (increased liver fat and levels of liver enzymes, and decreased low-density lipoprotein cholesterol and total cholesterol, including signals mapping to *PNPLA3* and *CILP2*). We detected highly significant enrichment of T2D association signals with evidence of ancestry-correlated allelic-effect heterogeneity (523 observed, 64.5 expected, binomial test $P<2.2x10^{-16}$), primarily driven by differences between East Asian and European ancestry GWAS. Association signals in the insulin production/processing cluster had larger effects on T2D in East Asian ancestry populations, whilst those in the obesity cluster had larger effects in European ancestry populations. These results highlight distinct biological pathways driving T2D susceptibility with heterogeneous effects across ancestry groups and emphasizes the importance of studying diverse populations to fully elucidate the pathogenesis of disease.

S16. Considering ancestry in study populations

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 082. Leveraging global multi-ancestry meta-analysis in the study of idiopathic pulmonary fibrosis genetics

Authors:

J. Partanen[1], P. Häppölä[1], W. Zhou[2,3], A. Lehisto[1], M. Ainola[4,5], E. Sutinen[4,5], R. Allen[6], A. Stockwell[7], O. Leavy[6], J. M. Oldham[8], B. Guillen-Guio[6], N. Cox[9], J. HIRBO[10], D. Schwartz[11], T. Fingerlin[11], C. Flores[12,13], I. Noth[14], B. Yaspan[7], R. G. Jenkins[15,16], L. Wain[6,17], S. Ripatti[1,4], M. Pirinen[4,1], T. Laitinen[18], R. Kaarteenaho[19], M. Myllärniemi[4,5], M. J. Daly[1,2,3], J. T. Koskela[1], International IPF Genetics Consortium, Global Biobank Meta-analysis Initiative (GBMI); [1]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Univ. of Helsinki, Helsinki, Finland, [5]Helsinki Univ. Hosp., Helsinki, Finland, [6]Univ. of Leicester, Leicester, United Kingdom, [7]Genentech, South San Francisco, CA, [8]Univ. of California, Davis, Sacramento, CA, [9]Vanderbilt Univ Med Ctr., Nashville, TN, [10]Vanderbilt Univ. Med. Ctr., Clarksville, TN, [11]Natl. Jewish Hlth., Denver, CO, [12]Hosp. Univ.rio N.S. de Candelaria, Santa Cruz de, Tenerife, Spain, [13]Inst. de Salud Carlos III, Madrid, Spain, [14]Univ. of Virginia, Charlottesville, VA, [15]Imperial Coll. London, London, United Kingdom, [16]Royal Brompton and Harefield Hosp., London, United Kingdom, [17]Natl. Inst. for Hlth.Res., Leicester, United Kingdom, [18]Tampere Univ. Hosp. and Univ. of Tampere, Tampere, Finland, [19]Univ. of Oulu, Oulu, Finland

Abstract:

Idiopathic pulmonary fibrosis (IPF) is a devastating fibrotic disease of the lungs - an orphan disease with disease rarity and poor prognosis causing challenges to research. GWAS of IPF, mainly restricted to European ancestry, have identified 23 loci highlighting genes involved in cell cycle and airway clearance among others. Extending the analyses to diverse ancestries has great potential to identify additional loci.

As part of the Global Biobank Meta-Analysis Initiative (GBMI), we present the first multi-ancestry IPF GWAS meta-analysis, incorporating 8,492 cases from electronic healthcare records (mostly based on PheCode 502) from 13 biobanks across the globe. We further meta-analyze GBMI with the latest IPF meta-analysis that included three clinically curated cohorts, reaching 11,160 cases and 1.36M controls representing six ancestries. Meta-analysis was performed using the inverse variance weighted fixed effects model.

With an over 4-fold increase in the number of IPF cases compared to the latest meta-analysis, we identified 25 genome-wide significant IPF loci out of which 7 were novel. Highlighting the importance of multi-ancestry meta-analysis, only one of the novel loci (*FUT6*) would have been identified if the analysis had been restricted to Europeans. One of the novel loci was East Asian specific (*PSKH1*), three (*GRP157*, *FKBP5*, and *RAPGEF2*) have been previously associated with lung function and genes at two loci (*GIPC2* and *FKBP5*) have been reported to be differentially expressed in IPF compared to control lung. Fine-mapping in FinnGen suggested deleterious coding causal variants at three loci (*KIF15*, *TERT*, and *SPDL1*), highlighting an unreported missense variant in *KIF15*. Effect size estimates were 2.1-fold (95%CI 1.8-2.5) higher for the clinical cohorts compared to the biobank studies. Additionally, we noted a 1.6-fold larger effect in males at *MUC5B*, the strongest genetic risk factor for IPF: OR_{males} (95%CI) = 3.2(2.9-3.6), $OR_{females}$ (95%CI) = 2.0(1.8-2.3), Cochran's Q p-value for heterogeneity = 3.4E-9, with consistent findings across biobanks but not replicating in four clinical cohorts. We also observed notable genetic overlap between IPF and COVID-19 hospitalization, with 7 of the 25 IPF loci reaching the FDR-adjusted nominal p-value in the COVID-19 analysis, out of which 3 were genome-wide significant. Genetic correlation between the traits estimated by LD-score regression was 0.35 (95%CI 0.14-0.56, p = 0.001).

In summary, we present the largest and first multi-ancestry IPF analysis further elucidating IPF genetics by both revealing novel loci and providing increased insight into previously identified ones.

S16. Considering ancestry in study populations

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 083. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries

Authors:

R. Liu[1,2], H. Gao[3], K. Song[4], Y. Kakuta[5], M. Li[6], H. Huang[1,2], Z. Liu[3]; [1]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [2]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Ctr. for IBD Res., Dept. of Gastroenterology, Shanghai Tenth People's Hosp. of Tongji Univ., Shanghai, China, [4]Dept. of Biochemistry and Molecular Biology, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [5]Div. of Gastroenterology, Tohoku Univ. Graduate Sch. of Med., Sendai, Japan, [6]Dept. of Gastroenterology, Third Affiliated Hosp. of Guangzhou Med. Univ., Guangzhou, China

Abstract:

Inflammatory bowel diseases (IBD) are a group of chronic disorders including Crohn's Disease (CD) and ulcerative colitis (UC). To date, most IBD genetic associations were derived in European ancestries, with a handful of non-European studies of much smaller scale, limiting the discovery and application of IBD genetics to the rest of the world population.

To address this issue, we conducted a large-scale IBD genetic study in East Asian populations (EAS) including subjects from China, Japan, and Korea, for a total sample size of 14,393 IBD (7,372 CD and 6,862 UC) and 15,456 controls, a 4x increase from the previous IBD EAS study. We identified 80 genetic loci associated with IBD beyond genome-wide significance, a substantial increase from 26 from previous studies. We found 26 coding variants in EAS, among which 13 have never been reported in EAS nor EUR. 11 of the 13 coding variants have higher minor allele frequency in EAS than in EUR. These EAS enriched coding variants implicated new IBD-associated loci (ADAP1 [P14R] and GIT2 [N387S]), directly pinpointed new IBD genes for the first time in known loci (CELA3B [R79W] and SHC1 [A205V]) and deepened the IBD allelic series in the same gene to facilitate target modulation in drug discoveries (PTAFR [A224D and N14S]).

We found that despite a cross-population genetic correlation close to one, polygenic risk models trained in EUR have reduced performance in EAS. Leveraging this sizable new EAS sample and a novel multi-ancestry PRS method, we constructed a PRS model that increased the IBD risk prediction accuracy in EAS by 2.4- and 1.8- fold for CD and UC, respectively. We have released this model to facilitate equitable deployment of genetic risk prediction.

Across EAS and EUR (FinnGen and non-Finnish Europeans), with a total sample size of 45,106 IBD cases and 353,562 control, we identified 81 new genetic loci associated with IBD increasing the total number of IBD-associated loci to 320. Our findings provided additional targets for modulation in drug discoveries through implicating known IBD pathogenesis pathways that have been identified as IBD therapeutic targets (e.g., IL-21 signaling). We also noted that IBD-associated loci have a strikingly large overlap with IgA nephropathy associated loci (10 out of 25 not counting MHC), suggesting the convergence of pathogenesis pathways for the two disorders that appear unrelated.

Our study represents a much-needed breakthrough in IBD genetics. Through the inclusion of the first sizable sample of non-European ancestry in IBD genetics, we made critical discoveries to uncover fundamental aspects of IBD pathogenesis that are relevant to East Asia populations and across the world.

S16. Considering ancestry in study populations

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 084. The extent of allelic heterogeneity in a trans-ancestry GWAS meta-analysis of alcohol and tobacco addiction in 3.4 million individuals

Authors:

G. Saunders, GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN), Trans-Omics for Precision Medicine (TOPMed); Univ. of Minnesota, Minneapolis, MN

Abstract:

The use and abuse of nicotine and alcohol account for >100 million disability-adjusted life years across the globe, constituting one of the world's leading public health problems. Despite this, the majority of genome-wide association studies thus far have been restricted to individuals of European ancestry, representing <1% of known worldwide genetic variation. Here, we leveraged a trans-ancestry GWAS of nicotine and alcohol use in up to 3.4 million individuals from 60 studies with recent ancestry from Africa (N = 119,589), America (N = 286,026), East Asia (N = 296,438), and Europe (N = 2,669,029). Overall, we identified 2,143 loci and 3,823 independent variants associated with our five substance use phenotypes: smoking initiation, age of initiation of regular smoking, cigarettes per day, smoking cessation, and alcoholic drinks per week. The trans-ancestry meta-analysis method allows for quantifying the extent to which associated variants differ in effect size by ancestry along four dimensions estimated from multi-dimensional scaling (MDS) of allele frequencies from each participating study. We found that 79.3% (N = 3,032) of independent variants did not differ in magnitude of effect sizes by ancestry. Of the remaining 791 variants, 136 (3.6% of all independent variants) showed strong evidence for allelic heterogeneity indicating that the effect sizes of these variants differ as a function of at least one axis of genetic variation. A single missense variant in the alcohol dehydrogenase gene *ADH1B* known to be protective against alcohol consumption showed effect size differences on three axes of ancestry variation. An increase on any of these three MDS components was associated with a reduced effect size of the protective allele, on average. Overall, we found that variants associated with alcohol and tobacco use have largely the same effects across population. This is consistent with the idea that the underlying genetic architecture of alcohol and tobacco use is similar across ancestry and informs our understanding of the reasons for reduced portability of polygenic risk scores across populations. While GWAS identified variants are not necessarily causal themselves, these results suggest that the generally low predictive accuracy of scores across populations that has been widely observed may be largely due to reasons other than difference in causal effect sizes, potentially highlighting the importance of differences in linkage disequilibrium patterns and allele frequencies.

S16. Considering ancestry in study populations

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 085. Epigenetic variation impacts ancestry-associated differences in the transcriptional response to influenza infection

Authors:

K. Aracena[1], Y-L. Lin[2], K. Luo[1], A. Pacis[3], S. Gona[2], Z. Mu[1], V. Yotova[4], R. Sindeaux[4], A. Pramatarova[5], M-M. Simon[5], X. Chen[6], C. Groza[7], D. Lougheed[3,8], R. Gregoire[3], D. Brownlee[3], Y. Li[1], X. He[1], D. Bujold[3,5], T. Pastinen[9,8], G. Bourque[3,5,8,6], L. Barreiro[1,2,10]; [1]Dept. of Human Genetics, Univ. of Chicago, Chicago, IL, [2]Dept. of Med., Univ. of Chicago, Chicago, IL, [3]Canadian Ctr. for Computational Genomics, McGill Univ., Montreal, QC, Canada, [4]Dept. of Genetics, CHU Sainte-Justine Res. Ctr., Montreal, QC, Canada, [5]McGill Genome Ctr., Montreal, QC, Canada, [6]Inst. for the Advanced Study of Human Biology (WPI-ASHBi) Kyoto Univ., Kyoto, Japan, [7]Quantitative Life Sci., McGill Univ., Montreal, QC, Canada, [8]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [9]Genomic Med. Ctr., Children's Mercy, Kansas City, MO, [10]Committee on Immunology, Univ. of Chicago, Chicago, IL

Abstract:

Humans display remarkable inter-individual variation in immune response when exposed to identical immune challenges. Yet, our understanding of the genetic and epigenetic factors contributing to such variation remains limited. Here we aimed to study the impact of epigenetic variation on ancestry-associated differences in immune response to influenza A (IAV) infection. To do so, we obtained monocyte-derived macrophages from 35 individuals with varying degrees of European- and African-ancestry. Following 24-hours of IAV infection, we collected from matched non-infected and infected samples data on gene expression (RNA-seq), chromatin accessibility (ATAC-seq), post-translational modifications of 4 histone marks (H3K27ac, H3K4me1, H3K4me3, and H3K27me3), and methylation levels, measured by whole genome bisulfite sequencing. In addition, we performed high-coverage (30X) whole genome sequencing of all individuals. We obtained over 211 billion reads across the different assays, generating the most extensive dataset to date of transcriptional and epigenetic variation in the response to IAV at the population level. Our results reveal that transcriptional and epigenetic changes in response to IAV infection are highly coordinated and likely driven by the activation of infection-induced transcription factors involved in the regulation of antiviral responses. We found that gene expression and H3K4me1 levels were the most divergent between ancestry groups- 37% of genes and 35% of H3K4me1 peaks tested showed a significant association with genetic ancestry levels- suggesting that ancestry-associated differences in gene expression are tightly coupled with variation in enhancer activity. Ancestry-associated differences in the epigenetic landscape are genetically controlled, even more so than variation in gene expression: up to 65% of the identified ancestry-differences in molecular traits are explained by cis quantitative trait loci (QTL) that are differentiated by genetic ancestry. Striking patterns of sharing across regulatory QTL indicate that genetic variation coordinately drives transcriptional and epigenetic inter-individual differences- on average, 57% of QTL identified in one data type in IAV-infected cells are shared with at least one other data type. Lastly, we show that among QTL variants that colocalized with immune-disease loci, only 7% were gene expression QTL, the remaining corresponding to genetic variants that impact one or more epigenetic marks, which stresses the importance of considering molecular phenotypes beyond gene expression in disease-focused studies.

S17. Genetic and functional advances in inherited neuromuscular disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 088. A whole exon targeted PCR/Nanopore sequencing assay that reveals SNVs, indels and CNVs across *SMN1* and *SMN2* with implications for SMA carriers and disease severity

Authors:

P. Rao, J. Kemppainen, N. Vishag, B. Martin, C. Fraher, G. Latham, B. Hall; Asuragen, a Bio-Techne Brand, Austin, TX

Abstract:

Introduction: Spinal Muscular Atrophy (SMA) is an autosomal recessive neuromuscular disease that results from pathogenic variants of the survival motor neuron 1 gene (*SMN1*). Copy number loss of exon 7 in *SMN1* gene is present in ~71-95% of carriers and 95% of SMA patients, whereas *SMN2* copy number can help inform SMA severity. Detection of silent carrier variants and additional disease-causing variants in *SMN1* has potential to improve the identification of carriers and those with SMA, particularly in underrepresented ancestral groups. The ACMG recommends a population neutral approach for SMA screening. *SMN1* and *SMN2* are paralogous genes with gene-specific variants that are challenging to resolve using short-read sequencing due to their high homology. Here we demonstrate a prototype long-read sequencing assay (LRS) on Oxford Nanopore Technologies' (ONT) MinION platform that accurately quantifies and differentiates *SMN1/2* copies (0 to ≥3) and intragenic variants in a single workflow. **Methods:** Cell lines (n=50; spanning 0 to ≥4 copies for both SMN1 and SMN2) were obtained from Coriell Cell Repository. Genomic DNA was isolated from whole blood of 200 presumed healthy donors. DNA was amplified, barcoded, pooled, prepped by ligation sequencing kit (ONT), and run on R9.4.1 flow cells using the Mk1B or Mk1C (ONT). Custom software was developed to manage projects, start sequencing, review run-time info, automate data analysis, and report genotypes. The AmplideX® PCR/CE *SMN*1/2 Plus kit (Asuragen) and Sanger sequencing were used as comparator methods to confirm copy number and other variants. **Results:** The prototype PCR/Nanopore assay accurately classified carrier status (i.e. 1+0 or 2+0 *SMN1* copies informed by silent carrier variants) in >97% of the samples and quantified 0 to ≥3 *SMN*1 and *SMN*2 copies consistent with comparator method in >95% genomic DNA samples. Point variants and indels such as c.*3+80T>G, c.*211_*212del in *SMN*1 and c.859G>C in *SMN*2 were detected. The assay detected deletions of exon7 and/or 8 in *SMN1* in all samples confirmed by comparator method. **Conclusions:** The data demonstrate that the prototype PCR/Nanopore assay can reliably detect CNVs, large exon deletions, SNVs and indels in the *SMN1/2* genes using a streamlined workflow. The assay addresses ACMG guidelines for standardized screening approaches and may also be combined with other common or challenging carrier screening genes to improve laboratory operations and economies in delivering more equitable carrier screening and diagnostic solutions. Whole-exon sequencing using LRS may also enable broader insights into *SMN1* pathogenic variants and *SMN2* disease modifiers across ancestries.

S17. Genetic and functional advances in inherited neuromuscular disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 089. *SPTSSA* variants alter sphingolipid synthesis and cause a complex form of hereditary spastic paraplegia

Authors:

X. Pan[1,2], S. Srivastava[3], H. Mor Shaked[4], K. Gable[5], S. D. Gupta[5], N. Somashekarappa[5], G. Han[5], P. Mohassel[6], M. Gotkine[4], E. Desroche[7], P. Goldenberg[8], Q. K. Tan[9], Y. Gong[10,11], B. Kleinstiver[12,11], C. A. Maguire[10], B. Wishart[8], H. Cope[9], C. Brito Pires[10,11], H. Stutzman[11,12], R. Spillmann[9], Undiagnosed Disease Network, R. Seyedsedjadi[10], O. Elpeleg[4], C-H. Lee[13], E. Simon[4], F. Eichler[10,11], T. M. Dunn[5], H. J. Bellen[1,2]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, [3]Dept. of Neurology, Rosamund Stone Zander Translational NeuroSci. Ctr., Boston Children's Hosp., Harvard Med. Sch., Boston, MA, [4]Dept. of Genetics, Hadassah Med. Ctr. and Faculty of Med., Hebrew Univ. of Jerusalem, Jerusalem, Israel, [5]Dept. of Biochemistry and Molecular Biology, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [6]Neuromuscular and Neurogenetic Disorders of Childhood Section, Natl. Inst. of Neurological Disorders and Stroke, NIH, Bethesda, MD, [7]Massachusetts Eye and Ear, Boston, MA, [8]Med. Genetics Unit, Dept. of Pediatrics, Massachusetts Gen. Hosp., Boston, MA, [9]Dept. of Pediatrics, Div. of Med. Genetics, Duke Univ. Sch. of Med., Durham, NC, [10]Dept. of Neurology, Massachusetts Gen. Hosp., Boston, MA, [11]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [12]Dept. of Pathology, Massachusetts Gen. Hosp., Boston, MA, [13]Dept. of Structural Biology, St. Jude Children's Res. Hosp., Memphis, TN

Abstract:

SPTSSA encodes an activating subunit of serine palmitoyltransferase (SPT), the enzyme that catalyzes the rate-limiting reaction of sphingolipid (SL) *de novo* synthesis. SLs are a diverse family of lipids with critical structural and signaling functions in the nervous system. The synthesis of SLs is tightly regulated and key to the homeostatic regulation are the ORMDL proteins which bind to SPT and mediate feedback inhibition of SPT enzymatic activity when SL levels become elevated. To date, no Mendelian disorder has been associated with *SPTSSA*. Through the Undiagnosed Diseases Network and international matchmaking, we identified three unrelated individuals with variants in *SPTSSA*. Two individuals have a *de novo* missense variant (c.152C>T, p.Thr51Ile), while one individual has a homozygous frameshift variant (c.171_172del, p.Gln58AlafsTer10). All three individuals present manifestations consistent with complex hereditary spastic paraplegia (HSP), including progressive motor impairment and spasticity, variable sensorineural hearing loss and language/cognitive dysfunction. The cryo-EM structure of the human SPT/ORMDL3 complex reveals that SPTSSA Thr51 resides in close association with ORMDL3, indicating its importance in the interaction between SPT and ORMDL3. Using biochemical and cell-based assays, we show that the p.Thr51Ile as well as the frameshift variant impair ORMDL regulation and cause excessive SL synthesis. Unrestrained SPT activity is evident from elevated SLs in serum and fibroblasts from the patients. In a fruit fly model we overexpressed human SPT and showed that it leads to excessive SL synthesis and causes severe motor defects and shortened lifespan, supporting the causal relationship between elevated SPT activity and some of the patient phenotypes. Moreover, while co-expression of human ORMDL3 reversed the elevated SL levels and rescued the phenotypes in flies expressing the wildtype SPT, it failed to rescue flies expressing the mutant SPT, confirming that the SPTSSA p.Thr51Ile variant abrogates ORMDL regulation of SPT *in vivo*. In summary, we show that variants of *SPTSSA* cause elevated SL synthesis and identify *SPTSSA* as the latest gene to be associated with complex HSP, a genetically and clinically diverse group of neurologic disorders.

S17. Genetic and functional advances in inherited neuromuscular disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 090. A reverse genetics and genomics approach to gene paralog function and disease: Myokymia and the juxtaparanode

Authors:

D. Marafi[1,2], N. Kozar[3], R. Duan[2], S. Bradley[3], K. Yokochi[4], F. Al Mutairi[5], N. W. Saadi[6], S. Whalen[7], T. Brunet[8], U. Kotzaeridou[9], D. Choukair[10], B. Keren[11], C. Nava[11], M. Kato[12], H. Arai[13], T. Froukh[14], E. A. Faqeih[15], A. M. AlAsmari[15], M. M. Saleh[15], F. Pinto e Vairo[16], P. N. Pichurin[16], E. W. Klee[16], C. Schmitz[16], C. Grochowski[2], T. Mitani[17,2], I. Herman[18,2], D. G. Calame[2], J. Fatih[2], H. Du[2], Z. Coban Akdemir[19,2], D. Pehlivan[2], S. Jhangiani[2], R. Gibbs[2], S. Miyatake[20], N. Matsumoto[20], L. J. Wagstaff[21], J. Posey[2], J. Lupski[2], D. Meijer[3], M. Wagner[22]; [1]Kuwait Univ., Jabriya, Kuwait, [2]Baylor Coll. of Med., Houston, TX, [3]Ctr. for Discovery Brain Sci., Univ. of Edinburgh, Edinburgh, United Kingdom, [4]4Dept. of Pediatrics, Toyohashi Municipal Hosp., Toyohashi, Aichi, Japan, [5]King Abdulaziz Med. City Riyadh, Riyadh, Saudi Arabia, [6]Coll. of Med., Univ. of Baghdad, Baghdad, Iraq, [7]Groupe Hosp.ier Pitié Salpêtrière, Paris, France, [8]Inst. of Human Genetics, Faculty of Med., Technical Univ. Munich, Munich, Germany, [9]Div. of Child Neurology and Inherited Metabolic Diseases, Ctr. for Pediatrics and Adolescent Med., Univ. Hosp. Heidelberg, Heidelberg, Germany, [10]Div. of Pediatric Endocrinology, Ctr. for Pediatrics and Adolescent Med., Univ. Hosp. Heidelberg, Heidelberg, Germany, [11]Département de génétique, Hôpital Pitié-Salpêtrière, Assistance Publique - Hôpitaux de Paris, 75013, Paris, France, [12]Showa Univ. Sch. of Med., Shinagawa-ku, Japan, [13]Dept. of Pediatric Neurology, Bobath Mem. Hosp., Osaka, Japan, [14]Philadelphia Univ., Amman, Jordan, [15]Section of Med. Genetics, King Fahad Med. City, Children's Specialist Hosp., Riyadh, Saudi Arabia, [16]Mayo Clinic, Rochester, MN, [17]Dept. of Pediatrics, Jichi Med. Univ., Yakushiji, Shimotsuke, Tochigi, Japan, [18]Dept. of NeuroSci., Boys Town, Natl. Res. Hosp., Omaha, NE, [19]UTHlth.Sch. of Publ. Hlth., Houston, TX, [20]Yokohama City Univ., Graduate Sch. of Med., Yokohama, Japan, [21]Ctr. for Regenerative Med., Inst. for Regeneration and Repair, Univ. of Edinburgh, Edinburgh, United Kingdom, [22]Inst. for Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany

Abstract:

INTRODUCTION The Leucine-rich glioma inactivated (LGI) family consists of four highly conserved paralogous genes, *LGI1-4*, that are highly expressed in mammalian central and/or peripheral nervous systems. LGI1 antibodies are detected in subjects with autoimmune limbic encephalitis and peripheral nerve hyperexcitability syndromes (PNHS) such as Isaacs and Morvan syndromes. Pathogenic variations of *LGI1* and *LGI4* are associated with neurological disorders as disease traits including familial temporal lobe epilepsy [MIM: 600512], and neurogenic arthrogryposis multiplex congenita 1 with myelin defects [MIM: 617468], respectively. No human disease has been reported for either *LGI2* or *LGI3*. METHODS We implemented exome sequencing and family-based genomics to identify cases with deleterious variants in *LGI3* and utilized GeneMatcher to connect practitioners and researchers worldwide to investigate the clinical and electrophysiological phenotype in affected subjects. We also generated *Lgi3* null mice and performed peripheral nerve dissection and immunohistochemistry to examine the juxtaparanode LGI3 microarchitecture. RESULTS We identified sixteen individuals from eight unrelated families with Loss-of-function (LoF) biallelic variants in *LGI3*. Deep phenotypic characterization showed *LGI3* LoF causes a potentially clinically-recognizable PNHS trait characterized by global developmental delay, intellectual disability, distal deformities with diminished reflexes, visible facial myokymia, and distinctive electromyographic features suggestive of motor nerve instability. Mouse studies revealed that LGI3 is highly expressed at the juxta-paranodal membrane and co-localizes with the voltage-gated potassium channels Kv1.1 and Kv1.2 and associated proteins. Moreover, loss of LGI3 results in reduced and mis-localized Kv1 channel complexes in myelinated peripheral axons. CONCLUSION Our data demonstrate biallelic LoF variants in *LGI3* cause a clinically-distinguishable disease trait of PNHS, likely caused by disturbed Kv1 channel distribution in absence of LGI3. Human paralogous gene mutational studies and aggregation of worldwide genomic and molecular data, of multiple variant allele types (SNV and CNV), from eight unrelated families with biallelic LoF variants in *LGI3* provide insights into: i) Clan Genomics, ii) organismal nervous system development and function, and iii) with mouse investigations informs the genesis of electrodiagnostic and clinically observed facial myokymia. Paralogous gene studies may also provide a route to molecular therapies.

S17. Genetic and functional advances in inherited neuromuscular disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 091. Predicting gene regulatory networks related to the development and maintenance of Schwann cells in the peripheral nervous system

Authors:

S. Khullar, J. Svaren, D. Wang; Univ. of Wisconsin - Madison, Madison, WI

Abstract:

Background One of the understudied nervous system cell types is Schwann cells, which associate with and myelinate axons in peripheral nerves. While Schwann cells perform similar roles to oligodendrocytes in the central nervous system, they have unique gene regulatory networks (GRNs) and capacities to foster nerve regeneration, potentially contributing to neuropathy. For example, studies of inherited peripheral neuropathies (Charcot-Marie-Tooth disease) identified over 100 disease genes, and transcription factor (TF) analysis has been used to identify target gene networks relevant to the disease. However, many genetic causes of neuropathy remain undefined. To provide resources to uncover noncoding variants emerging from the Whole-Genome Sequencing (WGS) analysis, we integrated emerging multi-omics data and predicted full gene regulatory networks for fetal and adult Schwann cells, linking regulatory elements with TF binding sites to target genes. Moreover, we identified genetic variants associated with such Schwann cell networks in tibial nerves and with related diseases.
Method: First, our analyses looked at Schwann cell chromatin accessibility data and found chromatin interactions between regulatory elements (e.g., promoters and enhancers). Then, we predicted TF binding sites on those interacting elements and linked TFs with binding sites on enhancers to target gene promoters into gene regulatory networks for both fetal and adult Schwann cells. Next, we used tibial nerve expression quantitative trait loci (eQTL) data, to predict noncoding variants linked to changes in target gene expression potentially via their disruptions of TF binding sites from our predicted regulatory networks. Further, we used recent WGS and Genome-Wide Association Studies (GWAS) data to find disease associations for those variants. Thus, we also included those disease variants into our networks via linking TF binding disruptions. Finally, we validated top TFs in our networks with peripheral nerve ChIP-seq data in rats. Results: The predicted gene regulatory networks consist of 693,707 and 842,845 TF-target gene edges for fetal and adult Schwann cells, respectively. Particularly, the fetal network links 423 TFs, 50,959 regulatory elements and 5,423 target genes, and the adult network links 423 TFs, 43,781 regulatory elements and 6,183 target genes. Across both networks, we found 651,991 shared TF-target gene edges (423 TFs, 5,349 target genes). Comparative network analysis also elucidated potential regulatory pathways (e.g. enhancers) dynamically altered during Schwann cell development and maintenance and in diseases (e.g. after peripheral nerve injury).

S17. Genetic and functional advances in inherited neuromuscular disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 092. A *C. elegans* model of neuropathy-associated *GARS1* mutations

Authors:

J. Pierluissi, A. Antonellis; Univ. of Michigan, Ann Arbor, MI

Abstract:

Charcot-Marie-Tooth disease (CMT) is a heritable peripheral neuropathy that is characterized by motor and sensory defects in the distal extremities. Mutations in five genes encoding aminoacyl-tRNA synthetases (ARSs)—a family of enzymes that ligate amino acids to cognate tRNA molecules—cause dominantly inherited CMT. One of these genes is glycyl-tRNA synthetase (*GARS1*), which encodes the enzyme responsible for ligating glycine to cognate tRNA molecules. There is wide allelic and clinical heterogeneity of *GARS1*-mediated neuropathy. We recently reported a 12-base-pair, in-frame deletion in *GARS1*, (E245_Q248; or ΔETAQ) in a patient with infantile-onset spinal muscular atrophy. The ΔETAQ mutation ablates enzyme activity *in vitro*, reduces viability in yeast complementation assays, and is dominantly toxic to mouse neurons. To determine the pathological significance of the ΔETAQ mutation and establish a pipeline to systematically study pathogenic *GARS1* variants, we employed a CRISPR/Cas9 method to generate a *C. elegans* model of *GARS*-mediated disease. Here, we will present data on the first *C. elegans GARS1*-mediated neuropathy model, which supports a loss-of-function effect of the patient variant. Heterozygosity for ΔETAQ *gars-1* produces a robust motility defect and fluorescent imaging of axons reveals neurotoxicity in the motor neurons. Pharmacological characterization of the neuromuscular junction of mutant worms indicates a degenerative defect in synaptic transmission, consistent with the patient phenotype. In addition to our unpublished data on this model, we will present plans to improve the phenotype observed in worm toward developing patient therapeutics. This work contributes to our understanding of the role of *GARS1* in peripheral neuropathy and establishes a framework for studying the pathogenicity of other ARS mutations of interest.

S17. Genetic and functional advances in inherited neuromuscular disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 093. Development of mini-gene therapy for Charcot-Marie-Tooth disease type 4B3 using patient-derived iPSCs

Authors:

E. Jacobs[1,2], C. K. Ramsoomair[1], M. Danzi[1], D. Isom[1], J. Medina[1], A. Rebelo[1], M. A. Saporta[1], S. Zuchner[1], M. Dohrn[1]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Med. Scientist Training Program, Univ. of Miami Miller Sch. of Med., Miami, FL

Abstract:

Charcot-Marie-Tooth Disease type 4B3 (CMT4B3) is rare, autosomal recessive hereditary neuropathy, presenting with severe symptom onset during infancy or early childhood. Patients have a wide clinical spectrum of symptoms ranging from an isolated demyelinating motor and sensory polyneuropathy to a complex neurodevelopmental phenotype with neuropathy, cranial nerve involvement, intellectual disability and facial dysmorphism. CMT4B3 is caused by a variety of mutations in the *Sbf1* gene, resulting in loss-of-function of the pseudo-phosphatase Myotubularin-Related Protein 5 (MTMR5). MTMR5 is an important regulator of endo-lysosomal trafficking and thus is involved in proper neuron and myelin development. Gene replacement therapy would be appropriate for CMT4B3 patients given the loss-of-function hypothesis, however the cDNA of *Sbf1* is larger (5,679bp) than the size limit of adeno-associated viral vectors (~4,700 bp). To circumvent this challenge, we've devised several candidate minigenes based on comparative protein family, cross species, and cross domain investigations. Additionally, we have also established a patient iPSC-derived human motor neuron system to both elucidate the mechanism of pathogenic alleles p.P1166TfsX5/+ and p.V1825GfsX27 on axonal degeneration and validate the mini-gene replacement strategy. Specifically, the proper subcellular localization, interaction with binding partners, and phenotypic rescue of the minigene are being explored in the iPSCs and other cell lines. Current functional studies into MTMR5 function include determination of motor neuron endo-lysosomal trafficking deficits, phosphoinositide metabolism defects and mitochondrial dysfunction. These important mechanistic findings will inform further refinement of candidate minigenes and other genetic therapies for CMT4B3.

S18. Genetic impacts on the epigenome and beyond

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 096. Enhancer cluster swapping reveals latent context sensitivity at the Igf2/H19 locus

Authors:

M. Maurano, W. Zhang, F. Zhu, G. Ellis, H. J. Ashe, A. Ribeiro-dos-Santos, R. Brosh, E. Huang, M. S. Hogan, J. D. Boeke, R. Ordoñez; NYU Sch. of Med., New York, NY

Abstract:

Enhancer function is frequently investigated piecemeal using truncated reporter assays or through genome editing of individual elements. Thus, it remains unclear to what extent regulatory elements can function independently of their native genomic context. To address this gap, we applied our Big-IN technology for targeted integration of large DNAs to dissect the regulatory architecture of the *Igf2*/*H19* locus, a paradigmatic model of enhancer selectivity. We assembled a series of 157-kb payloads containing a functional *Igf2/H19* locus and including mutations to genetically direct CTCF occupancy at the imprinting control region (ICR) which acts as a switch for the *H19* distal enhancer cluster. We delivered these payloads to scarlessly replace the endogenous locus in mouse embryonic stem cells (mESC), isolated and verified multiple independent clones for each, and differentiated them into mesendodermal cells for allele-specific expression analysis. Deletion analysis and delivery at a genomic safe harbor (*Hprt*) revealed a novel enhancer cluster outside the canonical *Igf2/H19* locus also capable of activating *H19* at long-range. This unexpected regulatory influence prompted us to investigate enhancer interchangeability more generally by exchanging components of the *Igf2/H19* locus with the *Sox2* locus, starting with the locus control region (LCR) essential for *Sox2* expression in mESC. Replacement of the *H19* enhancer cluster with the *Sox2* LCR was sufficient to confer robust expression of both *H19* and *Igf2* in mESC, which was partly responsive to CTCF occupancy at the ICR. Remarkably, the converse experiment showed that the *H19* enhancer cluster could not activate *Sox2* expression when relocated at the *Sox2* locus, but could partially induce *Sox2* at the *Igf2/H19* locus in an ICR-dependent manner, suggesting that unexpected dependencies may influence even the most studied functional elements. Our synthetic regulatory genomics approach permits large-scale manipulation of complete loci to understand how locus architecture relates to function.

S18. Genetic impacts on the epigenome and beyond

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 097. Genetic regulation of methylation across East Asian and European populations

Authors:

Y. Xia[1], T-T. Chen[2], S-C. Lin[2], C-Y. Chen[3], Y-C. Feng[4], T. Ge[5], H. Huang[6], Y-F. Lin[7]; [1]Broad Inst., Cambridge, MA, [23]Biogen, Cambridge, MA, [4]Natl. Taiwan Univ., Taipei, Taiwan, [5]Massachusetts Gen. Hosp., Boston, MA, [6]MGH, Boston, MA, [7]Natl. Hlth.Res. Inst.s, Taiwan, Zhunan Town, Miaoli County, Taiwan

Abstract:

DNA methylation (DNAm) quantitative trait loci (mQTL) reveal important knowledge of the noncoding genome to human health and disease. However, most mQTL studies were conducted in European populations (EUR), leaving uncertainties in the transferability of mQTL findings across populations and limiting our ability to interpret disease-associated variants. To alleviate this gap, we conducted the largest mQTL study to date in the East Asian populations (EAS), including 2,091 subjects of Han Chinese ancestry from the Taiwan Biobank using the Illumina EPIC array for methylation profiling, a two-fold increase from previous studies.

We tested the association between 8M genomic variants and variations in 772K DNAm sites. Using a stringent significance threshold ($p < 7e-14$), we identified 30M significant mQTLs, among which 568K were independent after 1000x permutations and conditional analysis. The mQTLs are enriched in DNase hypersensitivity regions (OR=1.13, $p<1e-22$) and enhancer regions (OR=2.01, $p<1e-22$).

Previous mQTL studies in EUR were almost conducted using the Illumina 450K array. So, more than half of the DNAm sites (53% out of 333K mQTL sites) were not interrogated. Compared with the largest mQTLs study so far in EUR with 10x more samples, we found that 64% of the mQTL sites we identified are novel, highlighting the importance of embracing ancestral diversity in studying the noncoding genome. In addition to the EPIC array specific DNAm sites, 9% of the mQTLs sites are EAS specific. These mQTLs have higher allele frequencies in EAS than EUR. For those shared in both populations, similar mQTL effect sizes (Two Sample t-test, $p = 0.13$) were found in both populations.

MQTLs provide a valuable resource to interpret the functional consequence of common variants underlying human complex traits and diseases in EAS, which has not been available until this study. We found that mQTL in EAS has the highest enrichment of heritability (h2) for SCZ in comparison with 20 other regulatory factors. The mQTLs SNPs were enriched with casual variants from the fine-mapping results of Biobank Japan and UK biobank, which may shed light on their functional mechanisms. Moreover, we found that mQTLs in EAS explained more h2 across traits in Biobank Japan than those in EUR (with a 10x larger sample size), suggesting the importance of using matched populations to interpret GWAS findings.

Taken together, we performed the largest mQTL study in EAS, with a comparison with EUR. We identified novel mQTLs, found similar effect sizes for shared mQTLs with EUR, and demonstrated that more GWAS h2 can be explained when using mQTLs derived from the same population.

S18. Genetic impacts on the epigenome and beyond

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 098. Genetic determinants of DNA methylation in African Americans: An meQTL mapping study in GENOA

Authors:

L. Shang[1], W. Zhao[1], Y. Wang[1], Z. Li[1], J. J. Choi[2], M. Kho[1], T. H. Mosley[3], S. L. R. Kardia[1], J. A. Smith[1], X. Zhou[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Wisconsin-Madison, Madison, WI, [3]Univ. of Mississippi Med. Ctr., Jackson, MS

Abstract:

Identifying genetic variants that are associated with variation in DNA methylation, an analysis commonly referred to as methylation quantitative trait locus (meQTL) mapping, is an important first step towards understanding the genetic architecture underlying epigenetic methylation variation. Most existing meQTL mapping studies have focused on individuals of European ancestry and are noticeably underrepresented in other populations, with a particular absence of large studies in populations with African ancestry. Lack of large-scale, well-powered meQTL mapping studies in populations with African ancestry can impede our understanding of the epigenetic mechanisms underlying gene expression and common diseases for these populations. We fill this critical knowledge gap by performing a large-scale in-depth cis-meQTL mapping study using the Illumina EPIC array in 961 African Americans from the Genetic Epidemiology Network of Arteriopathy (GENOA) study. We identified a total of 5,004,406 cis-acting meQTLs in 359,306 meCpGs after controlling for age, gender, population structure and familial relatedness. The CpG sites in identified cis-meQTLs are over-represented in the open sea, CpG island shore and shelf regions, as well as intergenic regions and gene bodies, but are under-represented in CpG islands as well as regions proximal to transcription start sites (TSS) of genes. Through conditional analysis, we found that 45.7% of meCpGs harbor multiple independent meQTLs, suggesting potentially polygenic genetic architecture underlying methylation variation. Through co-localization analysis, we found that a large percentage of the identified cis-meQTLs overlap with the expression quantitative-trait loci (cis-eQTL) identified from a previous study in the same population. Importantly, the identified independent cis-meQTLs explain a substantial proportion of cis-heritability in the associated meCpG sites (median = 85.86%), and, in mediation analysis, the cis-meQTLs also explain a substantial proportion (median=23.5%) of SNP heritability underlying gene expression. Overall, our results represent an important step toward revealing the causal/mediation role of methylation underlying gene expression, facilitating the functional integration and interpretation of epigenetic and gene regulatory changes underlying common diseases in African Americans.

S18. Genetic impacts on the epigenome and beyond

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 099. Identifying molecular mechanisms across common and complex neurological diseases by integrating single-cell epigenomics

Authors:

A. Marderstein, S. Kundu, A. Kundaje, S. B. Montgomery; Stanford Univ., Palo Alto, CA

Abstract:

Introduction: The majority of GWAS loci reside in non-coding regions. However, translating non-coding associations into mechanistic insights of disease is limited by incomplete knowledge of: (1) causal variants (due to linkage disequilibrium (LD)), (2) relevant cellular contexts, and (3) regulatory mechanisms linking SNPs to genes (which are cell-type-specific).
Methods: We used LD score regression to identify brain cell types enriched for GWAS heritability, integrating GWAS summary statistics from 10 neurological diseases with 70,631 scATAC-seq cells from 10 multi-region brain samples. We predicted the cell-type-specific effects of individual non-coding variants by training convolutional neural networks (called chromBPnet) that learn motif syntax and predict scATAC-seq profiles at single-base resolution from sequence. We used our chromBPnet models to nominate causal variants from a collection of 66,287 disease-linked SNPs (identified by either GWAS, colocalization with a GTEx brain eQTL, or LD with another variant of interest in the 1KG cohort), which complements existing fine-mapping methods limited by linkage disequilibrium.
Results: We established a single-cell resolution map of causal cell types across diseases, with GWAS heritability for 9/10 traits enriched within at least one cell-type-specific peak set. For example, while Alzheimer's heritability was enriched within only microglia peaks ($P<10^{-5}$), schizophrenia heritability was driven by SNPs within excitatory and inhibitory neuron peaks ($P<10^{-20}$) and was not enriched in microglia peaks. Using chromBPnet, we predicted candidate causal variants at 579 GWAS loci in specific cell types, such as an Alzheimer's SNP rs72962020 (near *PICALM*) disrupting a ELF5 motif in microglia cells. Our models prioritized multiple novel variants within difficult-to-resolve GWAS loci, such as 10 inhibitory neuron-specific SNPs at a recent Schizophrenia GWAS lead hit rs28490262 which contains 453 linked SNPs and is 1 Mb from the nearest gene. We show that cell-type-specific enrichment analyses of genes near prioritized variants captured cell-type-specific pathways missed by standard cell-type-agnostic pathway analyses, such as ion transport regulation in Alzheimer's disease via microglia-specific SNPs. Further, we found proximal high-scoring rare variants with higher effect sizes, which we are currently linking to multi-omic outlier effects.
Conclusion: Our results demonstrate that detailed maps outlining cell-type-specific regulation and computational models predicting the sensitivity of regulatory maps to variation enhance the mechanistic interpretation of non-coding loci.

S18. Genetic impacts on the epigenome and beyond

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 100. Targeted activation of the imprinted PWS locus via CRISPR/Cas9-based epigenome editing

Authors:

D. Rohm[1], J. B. Black[1], S. R. McCutcheon[1], A. Barrera[1], C. de Esch[2], D. J. C. Tai[2], M. Talkowski[2], N. Iglesias[1], C. A. Gersbach[1]; [1]Duke Univ., Durham, NC, [2]Massachusetts Gen. Hosp., Boston, MA

Abstract:

In mammalian diploid cells, imprinted genes retain a parent-of-origin specific epigenetic signature, resulting in monoallelic expression. One example of an imprinted locus is human chromosome 15q11-13, which is implicated in several distinct genetic disorders, including Prader-Willi Syndrome (PWS). PWS is a neurodevelopmental disorder caused by loss of expression of genes on paternally inherited chromosome 15, most commonly through a large deletion spanning 5-6 Mbp. In all cases, the maternally inherited allele is intact but epigenetically silenced. Thus, we sought to restore expression of the missing genes through targeted epigenetic modification of the silenced maternal allele using dCas9-based transcriptional activators. Although nuclease deactivated Cas9 (dCas9) fusions with transcriptional activators or repressors have been used to program epigenetic states, it is not clear to what extent and how these tools can reprogram more complex epigenetic regulation, such as imprinting. We first used CRISPR interference and activation screens to identify regulatory elements of the PWS gene *SNRPN* in human induced pluripotent stem cells (hiPSCs). To do so, we generated clonal reporter cell lines by specifically tagging endogenous paternal or maternal *SNRPN* with *2A-GFP* via genome editing. These high-throughput screens in the paternal and maternal *SNRPN-GFP* hiPSCs revealed regions within the PWS locus that abate paternal SNRPN-GFP expression or activate maternal SNRPN-GFP, respectively. We hypothesized that these regions may be involved in maintaining the epigenetic state of the PWS locus. Furthermore, we show that the transcriptional activator ^{VP64}dCas9^{VP64} and DNA demethylase ^{Tet1}dCas9 can each activate maternal *SNRPN* and downstream PWS transcripts, including the critical *SNORD116* cluster, in hiPSCs harboring a PWS deletion. Intriguingly, ^{VP64}dCas9^{VP64} and ^{Tet1}dCas9 function at unique regions and preferentially activate different *SNRPN* transcript variants. Using ATAC-seq and targeted bisulphite sequencing, we show that unlike ^{Tet1}dCas9, ^{VP64}dCas9^{VP64} increases chromatin accessibility without altering DNA methylation at the PWS imprinting center, indicating these two epigenome editors are functioning at the PWS locus through distinct mechanisms. We also find that transient delivery of ^{Tet1}dCas9 and a single gRNA leads to stable, long-term maternal *SNRPN* expression in hiPSCs. Our studies implement CRISPR screens to identify regulatory regions of an imprinted locus and offer insight into the epigenetic modifications that are sufficient to stably activate *SNRPN* and other PWS-associated transcripts in human cells.

S18. Genetic impacts on the epigenome and beyond

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 101. Single-sample, whole-genome parent-of-origin inference from imprinting control regions using ultra-long and Pore-C nanopore reads

Authors:

S. McKenzie[1], P. Rescheneder[1], S. Juul[2]; [1]Oxford Nanopore Technologies, Oxford, United Kingdom, [2]Oxford Nanopore Technologies, New York, NY

Abstract:

Parent-of-origin inference is a vital tool for clinical and research genetics, allowing the identification of compound heterozygous variants as well as variants with parent-specific phenotypic effects. It also allows researchers to link variants along a chromosome which might have differential interactions in cis vs. trans, and can simplify pedigree studies for candidate disease variants. Currently parent-of-origin inference is conducted almost exclusively by duo or trio sequencing, wherein one or both parents are sequenced along with a proband and parental variants are used for inference alongside local phasing information (from long reads, linkage disequilibrium, and/or population haplotype frequencies). However, obtaining sequencing data from parents is often logistically or financially difficult or prohibitive. Further, trio-based parent-of-origin inference can fail in genomic regions which are heterozygous for the same two haplotypes in the proband and both parents. Here we show that we can accurately assign parent-of-origin to the vast majority of heterozygous variants across the whole genome with a single-sample (i.e. only the proband) by using Oxford Nanopore Technologies (nanopore) ultra-long and Pore-C reads to simultaneously call and phase small nucleotide variants and 5mC CpG methylation across entire chromosomes. With ultra-long nanopore reads (reads with N50s greater than 50 kb), single nucleotide polymorphisms can be phased across tens of megabases, up to whole chromosome arms, while the addition of Pore-C reads allows the generation of accurate phase blocks which span entire chromosomes. Phased variants can then be used to phase the original nanopore reads and per-read methylation data, which allows the association of haplotypes with parent-of-origin specific methylation patterns at imprinting control regions. This method requires neither parental sequencing nor tissue-culture and can thus be performed economically on a single sample from a single proband. We validate our approach on the Genome in a Bottle reference HG002 and HG005 samples and compare against trio-binning with publicly available parental reads. We show high concordance for parent-of-origin inference between our method and the GIAB ground truth, comparable to trio-binning in both accuracy and completeness of parent-of-origin assignment.

S19. Genetics of human immunity, inflammation and infection

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 104. Dissecting IBD GWAS loci using single-cell and bulk multi-omics analysis

Authors:

Z. Xu[1,2], T. Wang[1], S. Feng[1,2], E. Heidrich-O'Hare[1], W. Chen[1], R. Duerr[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Tsinghua Univ., Beijing, China

Abstract:

INTRODUCTION: Several converging lines of genome-wide association study (GWAS) and mechanistic evidence implicate IL-23 signaling and the Th17 immune pathway in IBD pathogenesis. IL-23, IL-1β and prostaglandin E2 (PGE2) are overexpressed in human IBD tissues, modulate the development and function of human Th17 cells, and have receptors encoded by genes (*IL23R*, *IL1R1* and *PTGER4*) that map to IBD GWAS loci. **RATIONALE:** Bulk RNA-seq and ATAC-seq, and a single-cell assay for simultaneous transcriptome, cell surface protein epitope, and chromatin accessibility profiling (DOGMA-seq) enable a comprehensive and unbiased approach for fine mapping GWAS loci in disease-relevant cell types and conditions. **METHODS:** We used a large-scale multi-omics approach to dissect IBD loci in peripheral blood T cells that we activated and treated with IBD-relevant inflammatory mediators in short-term, *ex vivo* tissue cultures. T cells from seven human donors were activated and treated with four conditions (IL-1β and IL-23, plus either PGE2 or TGFβ, or both). A total of ~90,000 T cells were profiled using DOGMA-seq. $CD4^+CD45RO^+CD196^+$ T cells (an immunophenotype enriched for Th17 and regulatory T cells) from 38 human donors were activated and treated with IL-1β and IL-23, with or without PGE2, and assayed using bulk RNA-seq and ATAC-seq. **RESULTS:** We identified 17 T cell subtypes in DOGMA-seq data, including their activated and resting states. Cell-type specific regulator analysis revealed canonical and novel regulators for each subtype (*e.g., RORC* for Th17 cells and *ETS1* for resting states). Trajectory analysis of CD4+ naïve T cells corroborated regulators for activated/resting states. SNPs >95% likely to contain the causal variant in fine mapped IBD loci (IBD credible set SNPs) were enriched in Th1, Th17, MAIT, and regulatory T cell ATAC peaks. We observed fewer activated T cells and more resting T cells in conditions with PGE2, but TGFβ hardly influenced cellular composition. Among activated $CD4^+$ memory T cells, we observed fewer Th1 cells but more Th17 cells with PGE2 treatment. Differential gene expression analyses in Th1 and Th17 cells identified increased *CREM*, *PDE4D*, and *STAT4* and decreased *PTGER4* gene expression in conditions with PGE2. These findings were validated in bulk RNA-seq and ATAC-seq data in $CD4^+CD45RO^+CD196^+$ T cells. **CONCLUSION:** We mapped IBD credible set SNPs to specific T cell subtypes, highlighting future directions for candidate SNP functional studies. We found changes in T cell composition and identified differentially expressed genes upon PGE2 exposure in each T cell subtype.

S19. Genetics of human immunity, inflammation and infection

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 105. Methylation patterns associated with inflammation traits in racially and ethnically diverse populations

Authors:

J. Lundin[1], C. Kooperberg[1], U. Peters[1], Y. Hu[2], C. Carlson[1], J. Haessler[1], C. Haiman[3], K. North[4], L. Raffield[4], C. Avery[5], T. Matise[6], S. Buyske[6], C. Gignoux[7], E. Kenny[8], R. Loos[9], A. Reiner[10], S. Rich[11], J. Smith[12], N. Sotoodehnia[10], J. Bis[10], A. Lu[13], H. Xu[14], D. Levy[15], M. Cushman[16], The PAGE Study; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Cardio-CARE, Davos, Switzerland, [3]Univ. of Southern California, Los Angeles, CA, [4]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [5]Univ. of North Carolina, Chapel Hill, NC, [6]Rutgers Univ., Piscataway, NJ, [7]Univ. of Colorado Denver, Anschutz Med. Campus, Aurora, CO, [8]Icahn Sch. of Med. at Mt Sinai, New York, NY, [9]Icahn Sch. of Med. at Mt Sinai, New York City, NY, [10]Univ. of Washington, Seattle, WA, [11]Univ. of Virginia, Scottsville, VA, [12]Univ. of Michigan, Ann Arbor, MI, [13]Univ. of California, Los Angeles, Los Angeles, CA, [14]Univ. of Maryland, Baltimore, MD, [15]NHLBI/NIH, Framingham, MA, [16], Larner Coll. of Med. at the Univ. of Vermont, Colchester, VT

Abstract:

Chronic low-grade inflammation is part of a complex immune response that contributes to the pathophysiology of cardiovascular disease, cancer, and diabetes. In this study we evaluated DNA methylation associated with biomarkers of inflammation across multiple races and ethnicities. We also used genetic instruments to investigate the causal effect of DNA methylation on inflammation biomarkers to characterize the underlying genetic and epigenetic mechanisms.

We conducted and meta-analyzed epigenome-wide association studies of DNA methylation levels and two inflammation biomarkers C-reactive protein (CRP) and Interleukin-6 (IL-6) measured in peripheral whole blood among multi-ethnic discovery cohorts (ARIC, FHS, JHS, MEC, WHI) and diverse replication cohorts (AMISH, CHS, GENOA, and MESA). After quality control, we performed methylome-wide association analysis on ~420,000 CpG sites using mixed effects linear regression adjusted for age, sex, body-mass index, smoking, population structure, cell type composition (proportion of WBC species), study center (when multiple centers in a cohort), and technical covariates (as random effect). For CRP, CpG sites with $p < 1.1E-7$ (0.05/~420,000 CpG sites) in the discovery stage (n=10,995) were carried forward to an independent sample (n=3,116) for replication ($p<0.05$/number of sites). For IL-6, the analysis was only performed as a discovery set (n=4493; WHI, FHS, MESA, CHS). Two-sample summary-level Mendelian randomization (MR) analyses were performed for the replicated CpG sites associated with CRP to determine whether differential methylation was causal or consequential to the change in inflammation marker level.

We identified 992 CpG sites significantly associated with CRP values, and replicated 107 CpG sites at 88 loci (located more than 500kb away from other identified CpG site) in trans-race/ancestry and race/ancestry-combined meta-analysis. We identified 204 significant CpG sites at 161 loci associated with IL-6 in a cross-study analysis (discovery only). Using two-sample MR tests with inverse weighted regression methods, we identified 21 CpG sites that may mediate differences in in CRP levels.

Overall, we identified multiple novel inflammation-related CpG sites that provide new insights on the association of methylation and inflammation. On-going analyses will also include sensitivity analyses to evaluate the role of clinical comorbidities and other potential confounders. Further exploration of MR-Egger and weighted median models may assist in examining directional pleiotropy, testing heterogeneity, and understanding the direction of the causal pathways.

S19. Genetics of human immunity, inflammation and infection

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 106. Single-cell models reveal more pervasive T cell state-dependent regulatory effects than in bulk analyses

Authors:

A. Nathan[1], K. Ishigaki[1], L. Lecca[2], M. Murray[1], S. Raychaudhuri[1]; [1]Harvard Med. Sch./Brigham and Women's Hosp., Boston, MA, [2]Socios En Salud, Lima, Peru

Abstract:

Immune disease risk alleles from genome-wide association studies (GWAS) are enriched in non-coding regions. However, bulk T cell expression quantitative trait loci (eQTLs) have failed to explain most of these alleles' regulatory consequences. These variants may have cell-state-dependent effects that are not apparent in bulk analysis because the cell states that mediate disease may be more granular and thus missing in bulk assays of mixed cell states. Therefore, comprehensive characterization of the regulatory consequences of disease variants requires a single-cell model of cell states to define eQTLs with heterogeneous effects.Using a recently published Poisson framework for single-cell eQTL modeling (Nathan, et al. 2022 Nature), we systematically assessed the cell-state dependence of variants in immune disease GWAS loci in a multimodal CITE-seq dataset of 500,089 memory T cells from a Peruvian cohort. We defined high-resolution, functionally relevant, continuous cell states such as cytotoxicity and regulatory function by using canonical correlation analysis to integrate whole-transcriptome mRNA and 30 T cell surface proteins. Testing the associations between each gene and thousands of variants in a dataset of 500K cells is computationally expensive. To make the single-cell model more efficient without losing power through downsampling, we used within-individual neighborhood aggregation to reduce dataset size while maintaining structure. A ~5-fold reduction sped up the model ~5-fold, while effects were concordant with the full dataset (genotype z $r^2 = 0.97$, interaction z $r^2 = 0.89$).Focusing on 164 genes where eQTLs colocalize with rheumatoid arthritis GWAS, 54% had state-dependent lead eQTLs from pseudobulk T cell analysis. However, comprehensive analysis of all cis-variants—including those with weaker bulk effects and outside the promoter—found 68% had state-dependent effects. This meant that some eQTLs that weren't significant in the all-T cell analysis (e.g., *GSDMA*: rs12451100 beta = 0.13/p = 0.16) were stronger in certain cell states (beta = -1.82/p = $2.7x10^{-9}$ in cytotoxic CD4+ T cells) where they were even stronger than the lead effect (rs34170568 [lead] beta = 0.38/p = 0.20 in cytotoxic CD4+). eQTLs' state-dependence colocalized with GWAS signals in many loci (e.g., CD226, ORMDL3), and by estimating single-cell effect sizes from cells' states, we identified specific cells in which disease-associated eQTLs were strongest. These results suggest that many disease SNPs lacking bulk eQTLs may have cell-state specific effects, demonstrating the utility of single-cell models to decipher the regulatory impact of disease variants.

S19. Genetics of human immunity, inflammation and infection

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 107. Genetic and epigenetic regulation of immune cell subpopulations across health and ages

Authors:

M-J. Fave[1], E. Bader[1], M. Agbessi[1], E. Gbeha[1], V. Bruat[1], K. Skead[1], D. Soave[2], P. Awadalla[1]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Wilfrid Laurier Univ., Waterloo, ON, Canada

Abstract:

During aging, hematopoiesis is dysregulated leading to deleterious effects on health and contributes to increased risk for developing infections, cancer, chronic diseases, and death. However, multiple factors influence the aging process, from ancestry to environmental exposures and disease history, and are further confounded by age-associated diseases. Here, we profiled with multi-omics 400 participants selected from the extremes of the age and health risk spectrum to identify genetic, transcriptomic, and epigenomic features that contributes to healthy aging. We performed whole-genome sequencing, ATAC-seq and single-cell RNA-seq on over 500,000 cells and created a single-cell RNAseq reference dataset that consists of ~100K cells assigned to 22 clusters. We previously identified cell clusters associated with age and healthy aging, respectively NK and naïve B cells in males and CD4+T memory cells in females. Here, to further identify genes and pathways involved in healthy aging across multiple subtypes hematopoietic and immune cells, we mapped canonical and context-specific eQTLs across 22 cell clusters to identify genes and loci associated with health status and age. We identified cell types enriched in context-specific eQTL not previously mapped from pooled data, notably in CD4+T cells subtypes, suggesting the importance of genetic regulation at the cell-type level during healthy aging in adaptive cells. We also found context-specific eQTLs having opposite effects across different cell types. For example, an eQTL of *GIMAP4*, a GTPase of the immunity-associated protein known to be involved in T- and B-cell development and survival, shows opposite effect in B- and CD8+ T- cells subtypes in healthy versus non-healthy aged individuals. Finally, we performed multiple coinertia analyses across four age and health risk groups, which allow for the simultaneous detection of chromatin and expression features showing covariation in any of our 22 cell clusters. We reveal a stronger pattern of covariation between open chromatin profiles and gene expression in healthy aged individuals when compared to non-healthy agers, suggesting common protective epigenetic and genetic changes across cell types in healthy agers. Our results suggest that healthy agers might be able to delay or escape immunoscenescence through cell type specific regulation and decreased activation of stress and pro-oncogenic genetic regulatory mechanisms.

S19. Genetics of human immunity, inflammation and infection

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 108. Influenza GWAS of >1 million individuals identifies and experimentally validates two loci in *ST6GAL1* and *B3GALT5* and shows that risk factors for COVID-19 are largely distinct from flu

Authors:

J. Kosmicki[1], A. Marcketta[1], D. Sharma[1], A. Di Gioia[1], S. Batista[1], G. H. L. Roberts[2], J. B. Leader[3], H. Guturu[4], A. Verma[5], Regeneron Genetics Center, Penn Medicine Biobank, GHS-RGC DiscovEHR Collaboration, K. Siminovitch[6], D. Rader[7], J. Marchini[1], D. Carey[8], K. Rand[4], E. Hong[4], C. A. Ball[4], A. Baras[1], G. Abecasis[1], M. A. R. Ferreira[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Recursion Pharmaceuticals, SALT LAKE CITY, UT, [3]Geisinger Hlth.System, Danville, PA, [4]Ancestry DNA, San Francisco, CA, [5]Univ. of Pennsylvania, Philadelphia, PA, [6]Mount Sinai Hosp., Toronto, ON, Canada, [7]Univ of Pennsylvania, Philadelphia, PA, [8]Geisinger Clinic, Danville, PA

Abstract:

With few exceptions, genetic variants impacting risk for viral infections remain largely undefined. Even for one of the most common viruses, influenza, which infects 7% of the population every flu season, no genome wide association study (GWAS) has been published with more than 1000 samples. Given the success of the latest COVID-19 GWAS, coupled with the observation that SARS-CoV-2 and influenza viruses cause respiratory illnesses with shared symptoms and etiology, we sought to determine whether they share host genetic risk factors.

Utilizing self-reported data from the AncestryDNA cohort of individuals who consented to research (18,334 cases with a positive test for influenza and 276,295 controls), we found that common variants reported to associate with risk of SARS-CoV-2 infection or COVID-19 hospitalization in previous GWAS are not associated with risk of influenza infection, either individually or in aggregate. Consistent with these results, genome-wide genetic correlation (r_g) analysis showed that genetic risk factors for influenza are mostly distinct from those for SARS-CoV-2 infection (P=0.18; r_g=0.21; 95% CI -0.06-0.48), and only partially overlapping with those for COVID-19 hospitalization (P=0.004; r_g=0.39; 95% CI 0.12-0.66).

In the AncestryDNA cohort, we identified two genome-wide significant loci association with decreased risk of influenza infection and replicated both in ICD10-based cohorts totaling 30,339 cases and 1,095,953 controls: rs16861415 in intron 4 of *ST6GAL1* (P=1.4x10^{-10}, OR=0.86, 95% CI 0.83-0.90) and rs2837112 in the 3' UTR of *B3GALT5* (P = 1.3x10^{-19}, OR=0.90, 95% CI 0.88-0.92). Neither variant was associated with SARS-CoV-2 infection or COVID-19 hospitalization. The nearest genes to the influenza-associated loci were *ST6GAL1*, encoding an enzyme that catalyzes the transfer of sialic acid to galactose in the alpha-2,6 conformation and *B3GALT5*, encoding an enzyme that adds galactose to glycan structures. These genes are biologically relevant as sialic acid bonded to galactose in the alpha-2,6 conformation is an attachment factor for human influenza virus. We experimentally validated the *ST6GAL1* using an *in vitro* siRNA knockdown assay that resulted in ~50% lower influenza virus infectivity. Overall, our results show that genetic risk factors for COVID-19 and influenza are largely distinct and identify the first confirmed and experimentally validated GWAS loci for influenza.

S19. Genetics of human immunity, inflammation and infection

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 109. Novel genetic associations underlying symptom severity and infection susceptibility to SARS-CoV-2

Authors:

S. Andrews, The COVID-19 Host Genetics Initiative; Univ. of California, San Francisco, San Francisco, CA

Abstract:

The ongoing SARS-CoV-2 pandemic continues to pose a major public health threat especially in areas with low vaccination rates. The COVID-19 Host Genetic Initiative was formed to better understand the biological underpinnings of COVID-19 severity and SARS-CoV-2 infection, and has published GWAS meta-analyses to the community including the last release 6 (June 2021) of up to 125,584 cases and 2.5 million controls.
Here we present the latest (release 7) GWAS meta-analysis of up to 219,692 cases and over 3 million controls from 82 studies representing 35 countries, of which several are typically underrepresented in genetic studies. We performed a meta-analysis of three phenotypes: critical illness (respiratory support or death; 21,194 cases), hospitalization (49,033 cases), and lab-confirmed or self-reported PCR-confirmed SARS-Cov-2 infection (219,692 cases).
We found 30, 40, and 21, loci associated with critical illness, hospitalization, and infection due to SARS-Cov-2 respectively, for a total of 51 distinct genome-wide significant loci across all three phenotypes - adding 28 new genome-wide significant loci to the 23 previously identified. Using a two-class bayesian model for classifying loci, we identified 35 loci that are substantially more likely (>99% posterior probability) to impact disease severity (hospitalization) and 9 loci that influence susceptibility to SARS-CoV-2 infection.
Gene prioritization analysis highlighted three major biological pathways involved in susceptibility and severity. First, candidate causal genes located within loci associated with susceptibility to SARS-CoV-2 infection include ACE2, MUC5B, SFTPD, MUC16, MUC4, TMPRSS2, and SLC6A20, and are involved in viral entry pathways. Second, candidate causal genes located within loci associated with disease severity including TYK2, OSA1, IFNAR1/IFNAR2, JAK1, STAT1, and STAT2, play a role in the type I/III interferon pathway. Dysregulation of the immune response resulting from insufficient or delayed interferon response is observed in patients with severe COVID-19. Finally, several loci contain genes involved in the upkeep of healthy lung tissue and which have been previously associated with respiratory disease, lung cancer, or interstitial lung disease including DDP9, SFTPD, FOXP4, and MUC5B. In summary, this data release substantially expands the number of loci associated with COVID-19 severity and susceptibility and highlights pathways involved in the maintenance of the immune system and its regulation after viral exposure, and the upkeep of healthy lung tissue. The HGI's results are immediately made available at covid19hg.org with no restriction of use.

S20. High-throughput characterization of coding variants, from benchtop to desktop

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 112. Using Saturation Mutagenesis-Reinforced Functional Assays (SMuRF) to accelerate the pathogenicity identification of variants of uncertain significance

Authors:

K. Ma, S. Huang, N. Lake, K. Woodman, A. Lek, M. Lek; Yale Univ., New Haven, CT

Abstract:

The rapidly increasing list of variants of uncertain significance (VUS) discovered in individuals and our inability to interpret clinical consequences of these rare missense variants is an unappreciated challenge in the diagnosis of rare diseases. To address this urgent issue, we developed an adaptable high-throughput workflow called SMuRF (Saturation Mutagenesis-Reinforced Functional assays). SMuRF contains two components: 1) An accessible and adaptable cloning workflow to generate lentiviral plasmid library carrying all-possible small size variants, and 2) High throughput assay that characterizes the functional outcome of a variant directly correlated to the disease in contrast to measuring general properties such as protein stability or expression patterns. Dystroglycanopathies are caused by mutations of enzymes involved in the glycosylation of Alpha-dystroglycan (Alpha-DG). Here we report SMuRF's potential in characterizing all-possible single nucleotide variants of the FKRP enzyme and its impact on Alpha-DG glycosylation. We employed the FKRP lentivirus library to a perform a functional rescue in FKRP knock-out cell line. This rescue can be quantitated for each variant by using the IIH6C4 antibody that detects glycosylated Alpha-DG and can be used to separate cells by flow cyometry into high and low glycosylation groups, which enrichment can be quantitated by Next Generation Sequencing (NGS). We successfully generated functional scores for 4432/4455 (99%) of all possible *FKRP* single nucleotide variants (SNVs) and our result showed the expected trend for synonymous to have similar scores to wildtype and nonsense variants to be the most damaging. An analysis using ClinVar variants for FKRP was able to show 41/42 (98%) pathogenic and 140/150 (93%) benign variants have similar function scores to nonsense and synonymous variants, respectively. We are currently working on improving the versality of our SMuRF approach by 1) building variant plasmid library for other dystroglycanopathy genes, like *LARGE1*, and 2) developing flow cytometry-independent functional assays. We anticipate, with more adaptable components of the workflow, we can further expand applicability of SMuRF across different disease mechanisms in the hope to accelerate the interpretation of rare variants.

S20. High-throughput characterization of coding variants, from benchtop to desktop

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 113. Machine learning and genome editing to resolve variants of uncertain significance in *TSC2*

Authors:

C. Biar, J. D. Calhoun, G. L. Carvill; Northwestern Univ., Chicago, IL

Abstract:

Tuberous sclerosis (TS) is a multisystem mTORopathy characterized by benign tumors, drug-resistant epilepsy, and other cognitive manifestations, affecting 1 in 6,000 individuals born in the United States. This condition is caused by loss-of-function (LoF) genetic variants in the TSC1/2 complex, which result in constitutive mTOR signaling. As a result, ongoing clinical trials are testing mTOR inhibitors as a precision therapy for individuals with TS. However, access to precise treatment requires a precise genetic diagnosis. In *TSC2* alone, more than 2,220 variants of uncertain significance (VUSs) have been documented in ClinVar—and this number is likely to increase as next-generation sequencing becomes increasingly affordable and accessible. To address the growing challenge presented by VUSs, there is a critical need to develop tools to resolve their functional impacts. Our long-term aims are to (1) develop a *TSC2*-specific machine learning (ML) algorithm for variant pathogenicity prediction and (2) establish a high-throughput functional assay for *TSC2* VUS resolution. We have developed a ML model which utilizes ~20 features associated with variants in *TSC2*, including features related to evolutionary conservation and protein structure. We have also validated a functional assay whereby phosphorylation of S6 (P-S6), a well-characterized biomarker of mTOR pathway activity, distinguishes pathogenic from reference alleles. As a proof of principle, we used CRISPR/Cas9 genome editing to knockout *TSC2* in HAP1 cells. We then pooled $TSC2^{KO}$ and $TSC2^{WT}$ cells and tested whether FACS sorting by P-S6 level would be sufficient to enrich for cells with TSC2 LoF. Indeed, we observed that cells with high S6 phosphorylation were enriched for $TSC2^{KO}$ alleles. We next tested a known pathogenic missense variant, TSC2 p.Arg611Gln. Similarly, cells with constitutive mTOR signaling (high P-S6) were enriched for TSC2 p.Arg611Gln relative to unsorted cells or cells with low P-S6. Based on this and our previous work on mTORopathy-associated variants in *SZT2*, we conclude that sorting based on P-S6 labeling distinguishes LoF variants from WT. We will adapt this approach to incorporate prime editing-mediated saturation genome editing, increasing throughput to thousands of *TSC2* variants. This data will be used to test the validity of our ML pathogenicity predictions and to refine the performance of this classifier. This gene-specific workflow for improving the rate of VUS resolution is readily adapted to perform in other mTORopathy genes, such as *NPRL2*, *MTOR*, and *DEPDC5*.

S20. High-throughput characterization of coding variants, from benchtop to desktop

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 114. Functional classification of all possible SNVs in *VHL* using saturation genome editing

Authors:

G. Findlay, M. Buckley, N. Forrester, P. Dace; The Francis Crick Inst., London, United Kingdom

Abstract:

Germline mutations in *VHL* predispose individuals to various tumors, including renal cell carcinoma (RCC), hemangioblastoma, and pheochromocytoma. In ClinVar, over 700 variants of uncertain significance (VUS) have been reported in *VHL*, and hundreds of somatic variants have been encountered in tumor sequencing. Yet, the molecular mechanisms linking distinct VHL mutations to specific tumors have proven enigmatic. Saturation Genome Editing (SGE) uses multiplex CRISPR engineering to assay many human variants per experiment. Here, we applied a highly optimized SGE protocol to characterize nearly all possible single nucleotide variants (SNVs) in *VHL* coding regions. SNVs causing complete and partial *VHL* loss-of-function (LoF) were identified using next-generation sequencing and variants' effects on transcript abundance were measured with RNA-sequencing. In total, over 1,500 *VHL* SNVs were functionally scored, tiling nearly the complete coding sequence. The assay displays >95% sensitivity and 100% specificity for identification of pathogenic Type I VHL Disease variants. Sequence-function mapping delineates structurally critical regions of VHL intolerant to missense variation, nonsense variants in N- and C-terminal regions without deleterious effects, and dozens of splice-disruptive variants constituting a dosage-dependent gradient of functional impairment. Furthermore, 10% of VUS and 6% of variants absent from germline databases scored as LoF. We anticipate *VHL* SGE data will be immediately valuable for improving variant interpretation clinically. Furthermore, this work illustrates that systematic generation and analysis of functional data in light of structural information, human genetics data, and predictive models enables linking variants to the molecular mechanisms underlying cell-type specific phenotypes.

S20. High-throughput characterization of coding variants, from benchtop to desktop

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 115. A novel method to profile missense variation in extracellular proteins at scale reveals biochemical features important for expression and secretion of coagulation factor IX

Authors:

N. Popp[1], R. Powell[1], A. Chang[1], B. Zapp[1], K. Lannert[2], J. Johnsen[3,2,4], D. Fowler[1,5]; [1]Univ. of Washington, Seattle, WA, [2]Bloodworks Res. Inst., Seattle, WA, [3]Univ. of Washington Sch. of Med., Seattle, WA, [4]Washington Ctr. for Bleeding Disorders, Seattle, WA, [5]Brotman Baty Inst. for Precision Med., Seattle, WA

Abstract:

Multiplexed Assays of Variant Effect (MAVEs) are a powerful tool for assessing the effects of variation in proteins to understand sequence-function relationships, inform biological structures, and infer pathogenicity in clinical settings. Currently, the genes that can be studied with MAVEs are limited to those which encode for intracellular or membrane-bound proteins, because functional selection on the protein and the subsequent genomic DNA sequencing readout for identifying variants must be spatially connected within an individual cell. We hypothesized that cell surface display, as has been used in yeast and phage, could be adapted for use with MAVEs in mammalian cells. By fusing an extracellular protein to a single pass transmembrane domain, protein variants can be displayed on the surface of mammalian cells and assayed for function, thus re-establishing the spatial connection between functional selection and genomic DNA sequencing readout necessary for MAVEs.
As proof of concept for our mammalian cell-surface display system, we first assayed for expression of coagulation factor IX (FIX), a secreted plasma protein. Variation in the *F9* gene can cause decreased levels of circulating FIX, leading to the coagulation disorder hemophilia B. We applied this technique to a library of nearly all missense variants in FIX to profile each variant's effect on FIX secretion. We find that 37.7% of missense variants are poorly secreted in our system. Furthermore, our secretion assay alone explains half of the variance in clinical FIX antigen levels from a multinational database of patients with hemophilia B. We further find that cysteine substitutions have uncharacteristically strong negative effects on FIX expression, supporting an important role for disulfide bonds and redox conditions in the production and expression of FIX. These findings contrast with data from cytoplasmic proteins where cysteine substitution is generally well-tolerated. Lastly, we expand our cell surface display system to other proteins, showing the generalizability of our method for assessing variation in secreted proteins at scale.

S20. High-throughput characterization of coding variants, from benchtop to desktop

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 116. The landscape of regional missense mutational intolerance quantified from 125,748 exomes

Authors:

L. Wang[1,2,3], K. R. Chao[1,2], Genome Aggregation Database Consortium, A. H. O'Donnell-Luria[1,2,3,4], H. L. Rehm[1,2,3], D. G. MacArthur[2,3,5,6], M. E. Talkowski[1,2,3], G. Tiao[1,2], K. J. Karczewski[1,2], M. J. Daly[1,2,3], K. E. Samocha[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Harvard Med. Sch., Boston, MA, [4]Boston Children's Hosp., Boston, MA, [5]Garvan Inst. of Med. Res., and UNSW Sydney, Sydney, Australia, [6]Murdoch Children's Res. Inst., Melbourne, Australia

Abstract:

Missense variants have been broadly established to play a key role in common and rare human diseases. However, interpreting these variants remains challenging, as their functional consequences are highly variable depending on their location and amino acid substitution. The initial derivation of the missense badness, PolyPhen-2, and constraint (MPC) score from ExAC data sought to integrate information about subgenic regional missense constraint (RMC) and variant-level metrics to predict missense deleteriousness. While these metrics have been employed across massive disease association efforts, their resolution remained limited, primarily due to the size of the ExAC dataset (N = 60,706). Recently developed population references of much greater size and diversity now offer increased power to strengthen the precision of these metrics and elevate their utility. Here, we leverage the power of 125,478 exomes in gnomAD v2.1 to update the MPC and RMC metrics for broad application in association studies and variant interpretation. Major method refinements improve the model of expected missense variation and introduce per-base resolution of constrained region breakpoints. The updated metrics will be available in the gnomAD browser, and the underlying codebase has been released in an open-source GitHub repository. Using the updated metrics, we discover 3,655 canonical gene transcripts that harbor regional differences in missense constraint, 52.3% of which had not been detected with ExAC. Genic regions predicted to be highly constrained (observed/expected missense fraction < 0.4) are found to align more closely with protein domains. Initial analyses reveal a 10-fold enrichment of *de novo* missense mutations predicted to be highly deleterious (MPC ≥ 3) in 37,488 individuals with neurodevelopmental disorders compared to unaffected individuals, and an increase in predicted deleteriousness percentile of pathogenic haploinsufficient ClinVar missense variants (Wilcoxon $p = 0.001$) consistent with their putative phenotypic impact. Further expansion of population reference sets will continue to enhance the accuracy with which constrained regions can be defined, which is supported by the observation that transcripts with more expected missense variants are predicted to have more regional differences in missense constraint (Pearson's $r = 0.30$, $p < 10^{-16}$). The improved metrics described here provide a finer resolution on the landscape of missense constraint across the coding genome, and we expect they will be of value to clinical variant interpretation and gene discovery efforts in human disease.

S20. High-throughput characterization of coding variants, from benchtop to desktop

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 8:30 am - 10:00 am

ProgNbr 117. Prediction of recessive inheritance for missense variants in human disease

Authors:

B. Petrazzini[1], D. Balick[2], I. Forrest[1], J. Cho[1], G. Rocheleau[1], D. Jordan[1], R. Do[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Harvard Med. Sch., Boston, MA

Abstract:

Current variant prediction methods are designed to discriminate pathogenic and benign missense variants. However, further axes of information are needed to achieve a holistic representation of the genotype-to-phenotype map. For instance, basic genetic attributes such as mode of inheritance (MOI) information are missing from current annotation pipelines. Recent attempts to develop MOI prediction tools show moderate discrimination performance, lack robust validation and are insensitive to recessive-acting mechanisms. Recessive variants are particularly hard to identify due to the scarcity of family pedigree data, making computational tools an important asset to tackle recessive-acting diseases in clinical practice. The lack of clinically useful MOI prediction tools highlights a need for robust recessive variant identification techniques.Here we present MOI-Pred, a three-way variant-level MOI prediction tool that labels missense variants as pathogenic for autosomal recessive (AR) disease, pathogenic for autosomal dominant (AD) disease, or benign. MOI-Pred fits a random forest algorithm to 1,248 dominant and 2,481 recessive pathogenic variants from ExoVar, together with 3,729 presumably benign variants from gnomAD. These were annotated with 78 features providing evolutionary, functional and allele frequency information. Predictions from the three-way classification model were validated using 255 dominant and 261 recessive pathogenic variants from ClinVar, together with 1,010 presumably benign variants from GEM not seen by MOI-Pred nor any of its constituent scores, ensuring reliable performance metrices in novel variation. MOI-Pred shows strong ability to discriminate variants across the three classes with AUROC=0.99/0.99/0.96, sensitivity 0.85/0.87/0.93, and specificity 0.98/0.99/0.86 for AR/AD/benign classes. Additionally, validation of the MOI-Pred predictions using real world electronic health record data shows MOI-Pred recessive predictions are enriched for recessive associations with human diseases (odds ratio = 4.30, 95% CI=4.07 to 4.55) in ClinVar Pathogenic variants. Finally, we demonstrate utility of MOI-Pred for clinical assessment of individual variants. Single variant association testing of predicted AR variants from MOI-Pred identifies three recessive associations for clinical diseases in the EHR-linked Bio*Me* Biobank. To date, MOI-Pred is the only approach that can predict MOI with high discrimination, robust validation and demonstrated clinical utility. MOI labels for 71M human missense variants can be found at https://github.com/rondolab/MOI-Pred/.

S21. Averting Alzheimer's as soon as possible

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 120. Single-cell transcriptomic and epigenomic dissection of Alzheimer's disease pinpoints causal GWAS variants and reveals epigenome erosion

Authors:

X. Xiong[1], C. Boix[1], **B. James**[2], L. Hou[3], K. Galani[1], L-L. Ho[1], Y. Park[4], N. Sun[1], H. Mathys[1], D. Bennett[5], L-H. Tsai[1], M. Kellis[1]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]MIT & Harvard, Cambridge, MA, [3]MIT & Broad Inst., Cambridge, MA, [4]Univ. of British Columbia, Vancouver, BC, Canada, [5]Rush Univ. Med. Ctr., Chicago, IL

Abstract:

Genetic studies of Alzheimer’s disease (AD) revealed multiple associated loci, providing hope for new therapeutic insights. However, the vast majority of AD-associated variants lie outside protein-coding regions, making it difficult to recognize driver variants, target genes, and the cell types where they act. Interrogating the epigenomic landscape and regulatory circuit of healthy and AD individuals in a cell-type specific manner can help pinpoint the GWAS variants and establish the mechanistic links between these variants and disease etiology. Here, we profile single-cell ATAC-seq and single-cell RNA-seq from the prefrontal cortex of 92 individuals at various stages of AD progression (29 early-AD, 15 late-AD, 48 age-matched controls). We capture 1M cells after QC, which we annotate into 11 subtypes of excitatory neurons, 13 subtypes of inhibitory neurons, microglia, oligodendrocyte, astrocyte, OPC, and vascular cells. We show that scATAC peaks in microglia strongly and specifically enrich for AD GWAS signal, concentrated in annotated enhancers showing H3K27ac and lacking H3K4me3. We integrate scATAC and scRNA data to link non-coding variants to candidate target genes through enhancer-gene linking, across multiple lines of evidence, including: (1) cell-type-specific and AD-differential accessibility; (2) scATAC QTLs; (3) QTL-GWAS colocalization; (4) transcription factor binding enrichment; and (5) gene-peak linking with ATAC co-accessibility, HiChIP, EpiMap, and eQTL information. These analyses result in high-confidence fine-mapped variants, including for well-known AD driver genes BIN1 and PICALM.Late-stage AD samples showed a striking global change in their chromatin accessibility landscape indicative of genome-wide “epigenome erosion”, whereby accessible regions close and inaccessible regions open, indicative of loss of cell type identity. These changes were also reflected in our matched transcriptional profiles, with cells from epigenome-eroded individuals strongly enriched for transcriptionally-deidentified cells whose expression patterns lie between other cell types, particularly for excitatory neurons and oligodendrocytes. Overall, our study provides a systematic comparison between the epigenomic landscape of healthy and AD cohorts at single-cell resolution, pinpoints candidate AD-driver variants and genes, and reveals global changes indicative of late-stage dysregulation.

S21. Averting Alzheimer's as soon as possible

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 121. Genome-wide circRNA dysregulation contributes to Alzheimer's disease pathogenesis

Authors:

Y. Li[1], F. Wang[1], H. Shen[1], P. Martinez-Feduchi[1], Y. Feng[2], B. Yao[1]; [1]Emory Univ. Sch. of Med., Dept. of Human Genetics, Atlanta, GA, [2]Emory Univ. Sch. of Med., Dept. of Pharmacology and Chemical Biology, Atlanta, GA

Abstract:

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most prevalent cause of dementia with no curative treatments available to date. While major efforts focus on characterizing the functions of AD risk coding genes, the contribution of non-coding RNAs in AD pathogenesis remains largely unknown. Circular RNAs (circRNAs) are a novel class of covalently closed non-coding RNAs that are highly enriched in the brain and stable and accumulable during aging progression. Recent data established a strong link between circRNA expression and AD onset, progression and severity. However, mechanism and dynamic regulation of circRNA landscape prior and after early AD onset is not investigated. Our recent work established a robust and sensitive platform to identify genome-wide circRNA landscape combining published A-tailing RNase R approach with our computational framework CARP (CircRNA identification using A-tailing RNase R approach and Pseudo-reference alignment). Using A-tailing and CARP, we generated accurate and comprehensive cortical circRNA landscape and dynamic regulations from 5xFAD mouse model and their littermate controls spanning several aging points. We identified circRNA dysregulation in critical time window of AD early onset and further classified them into different groups according to their dynamic expression patterns during aging. We explored the potential biological roles of key circRNA associated with AD in interfering with miRNAs and RNA-binding proteins. In particular, we found a circRNA that highly conserved between mouse and human and could sponge multiple AD related miRNAs to regulate expression of their targets, which are significantly functional enriched in neurogenesis and neuron differentiation. Importantly, the relatively low expressed circRNAs that could specifically be detected by A-tailing method also contribute to miRNA sponge through additive effect. At the circRNA biogenesis level, circRNAs that dysregulated in AD were regulated by either *trans* regulatory element RBPs that bind to their flanking intron and *cis* element A-to-I editing in the reverse complementary elements in their flanking intron. Together, our advanced experimental and computational method identified conserved dysregulated circRNA in early AD onset as well as molecular mechanism of their function and biogenesis.

S21. Averting Alzheimer's as soon as possible

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 122. Spatial transcriptomics based gene interactions analysis in AD brain

Authors:

S. Wang, J. Greenbaum, H-W. Deng; Tulane Univ., New Orleans, LA

Abstract:

Although previous studies have explored the spatial gene expression profiles of Alzheimer's Disease (AD) brain by spatial sequencing experiments, the annotation of brain region, phenotype (AD and wild-type (WT)) and Aβ index information allowed the research for unique genome-wide transcriptomic association features in different spatial and pathological states. However, the dynamic gene interactions on different aspects during Aβ accumulation remain largely unknown.
In this study, we performed ligand-receptor (L-R) communication, Transcription factor (TF) regulatory network and spot-specific network (SSN) analyses to reveal gene associations based on a spatial transcriptomics dataset of App knock-in mouse mode. In general, our work revealed 17 L-R pairs with contrary tendencies through Aβ accumulation process and showed the specific L-R interactions across the whole brain areas or hippocampus layer in different extent of pathological change. To explore the function of unique L-R pairs that are only differentially expressed in one region pair (whole brain areas or hippocampus (HP)), we searched their scores in the AlzData database. Whole brain unique L-R pair gene CNR2, IL15, DLL1 in 3w group, ROR2, C3, CXCL16 in 18w group and HP layer unique L-R pair gene C3, FGFR2, CNR2 in 3w group, FLT1 in 18w group got the highest convergent functional genomic (CFG) score, revealed the potential of unique L-R pairs' AD related function. We also identified nerve function related TFs in HP and ENTI. A gradually up-regulated regulon cluster (cluster A) in HP was involved in negative regulation of humoral immune response and DNA-binding transcription activator activity, RNA polymerase II-specific GO terms. This dynamic change showed cluster A regulons may be related to the humoral immune response and the catastrophic failure of transcription-related molecules (especially RNA polymerase II)' transport between the cytoplasm and the nucleus in late AD neurotoxic process. GO analysis results of 32 active regulons in the L1 and L2 group of early affected ENTI area revealed their wide influences at early AD stage including neurotransmitter, synaptic and cognition related functions. We then calculated different network degree matrix (NDM) value genes with different transcriptomic interactions state to reveal unique gene connections in certain phenotype (WT/AD) of different brain regions and age groups.
This is the first study to identify the gene associations through Aβ accumulation based on spatial transcriptomics and establishes the foundations to reveal advanced mechanisms of AD in a new perspective based on the spatio-temporal comprehensive gene interactions.

S21. Averting Alzheimer's as soon as possible

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 123. Large-scale multi-omic analyses in CSF identified multiple causal and druggable targets for Alzheimer's disease

Authors:

C. Cruchaga[1], D. Western[1], C. Wang[1], J. Tismina[1], R. Agustin[1], L. Wang[2], P. Pastor[3], Y. Sung[1]; [1]Washington Univ., Saint Louis, MO, [2]Washington Univ., St Louis, MO, [3]FIMA Ctr. for Applied Med. Res., Pamplona, Navarra, Spain

Abstract:

Expression quantitative trait locus (eQTL) studies have been instrumental in pinpointing the functional genes and identifying drug targets for many complex diseases. However, we and others have demonstrated that QTLs using other omic layers (proteins, metabolites, and lipids) do not overlap with eQTL and provide a biological context for additional GWAS loci. Cerebrospinal fluid (CSF) biomarkers are one of primary diagnostic tools in Alzheimer's disease (AD), highlighting CSF's relevance to brain aging and AD pathology. There were protein (pQTL) and metabolite QTL (metabQTL) studies using CSF, but only in small sample sizes. Here, we present a large-scale CSF pQTL and metabQTL analysis of over 3,000 individuals. We generated proteomics (Somalogic; 7,584 proteins) and metabolomics (Metabolon; 440 analytes) data for 3,065 individuals in CSF. After rigorous quality control, we performed QTL analysis using genomic data imputed using HG38 TOPMed reference panel, in three stages: discovery, replication, and meta-analyses. We performed colocalization of our QTLs with GWAS for AD risk. Mendelian randomization (MR), protein-wide and metabolite-wide association study (PWAS/MWAS) using FUSION was performed to identify proteins and metabolites that are causal for AD. In our proteomic analysis, we identified a total of 2,472 significant pQTL (1,297 in cis and 1,175 in trans), of which 1,339 are novel. Of the 99 GWAS loci for AD risk, 68 had a pQTL with suggestive significance (1×10^{-5}) and 36 with genome-wide significance (5×10^{-8}). PWAS analyses identified over 25 proteins causal for AD that are enriched in multiple pathways including amyloid beta metabolism (APOE, ACE and CNTN2), endolysosomal (GRN, CTSH, CLN5 among others) and immune pathway (TREM2, CD33, IL34 among others). In our metabolomic analysis, we identified 192 (113 novel) metabQTL. Of these, 16 metabQTL colocalized with AD risk loci. These metabolites are enriched on cortisol and sphingomyelinase cholesterol pathways. This study represents the largest QTL analysis of CSF to date and identified hundreds of novel pQTL and metabQTL. Many of our CSF pQTLs were distinctive from recently published plasma pQTLs, indicating the presence of many tissue-specific signals. We also identified several causal and druggable proteins and metabolites for AD. This highlights a need for QTL studies in additional tissues including CSF, a good proxy for brain tissue. Our findings elucidate the pathological events that lead to AD, which can provide critical insights for clinically translatable interventions for prevention and treatment.

S21. Averting Alzheimer's as soon as possible

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 124. Rare-variant analysis of whole-genome sequence data obtained from multi-ancestry families identifies new genes associated with late-onset Alzheimer's disease

Authors:

S. Leal, Y. Liu, C. Li, A. Lee, B. Vardarajan, G. Wang; Columbia Univ., New York, NY

Abstract:

Late-onset Alzheimer's disease (LOAD) is the most common form of AD. Existing literature suggests AD is genetically heterogeneous with a heritability of 60-80%. Only a small fraction of AD heritability can be attributed to common genetic variants that were identified through population-based genome-wide association studies (GWAS). Rare variants (RV) also likely contribute to AD heritability. Here we report on a family-based, multi-ancestry RV association study for LOAD using whole-genome sequence (WGS) data. We analyzed 1,068 families from the *Estudio Familiar de Influencia Genetica en Alzheimer* (EFIGA) and NIA-LOAD cohorts with N=3,479 members with WGS data [313 European families (N=1,042), 553 Caribbean Hispanics families (N=2,133), and 198 African American families (N=288)]. We also analyzed case-control exome sequence data from the Alzheimer's Disease Sequencing Project (ADSP), consisting of 2,142 Caribbean Hispanics, 3,988 African Americans, and 11,123 individuals of European ancestry. Generalized linear mixed models were used to implement gene-based RV aggregate association tests via the SMMAT software, Gene level meta-analyses were performed across ancestries. For the multi-ancestry LOAD families, we identified 10 associated genes [p-value < 2.8×10^{-6} (Bonferroni correction for testing ~18,000 genes)]. Of the 10 identified genes, *TNFRSF21*, *EML6,* and *AHI1* are differentially expressed genes (DEG) for LOAD pathology in single-cell data, and *TNFRSF21*, *CACNG7*, *RAP2B*, *AHI1,* and *SPRED1* are DEG in bulk RNA-seq data. These genes, along with *RAP2B, HSF2,* and *POGZ*, have previously been reported to be involved in AD-related neurodegenerative and neuropsychiatric disorders but were not identified in GWAS studies. Meta-analysis of the families with the case-control ADSP data revealed several additional genes, including *KCNJ15* which was previously reported to be associated with AD in a Chinese cohort common variant GWAS. Additionally, several previously published LOAD GWAS genes, including *ADAMTS16*, *ABI3, ADAMTS12*, *AMMECR1L*, *APH1B, CABIN1*, *CLUH*, *EPDR1*, and *SORL1* showed suggestive evidence of association (p-value < 0.01) in our RV aggregate analysis, suggesting that common and rare variants in the same gene may confer LOAD risk. Our work also demonstrates the important role a multi-ancestry family-based design plays in the study of complex traits.

S21. Averting Alzheimer's as soon as possible

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 125. Multi-ancestry genome-wide association analysis of late-onset Alzheimer's disease (LOAD) in 60,941 individuals identifies novel cross-ancestry loci

Authors:

A. Naj[1], C. Reitz[2], F. Rajabli[3], G. Jun[4], P. Benchek[5], G. Tosto[2], j. sha[1], C. Zhu[4], N. Kushch[3], W-P. Lee[1], J. Haut[1], K. Hamilton-Nelson[3], N. Wheeler[5], Y. Zhao[1], J. Farrell[4], J. Chung[4], M. Grunin[5], Y. Leung[1], D. Li[4], E. Lucio da Fonseca[3], J. Mez[4], E. Palmer[5], J. Pillai[6], R. Sherva[4], Y. Song[5], X. Zhang[4], T. Iqbal[1], O. Pathak[1], O. Valladares[1], A. Kuzma[1], B. Kunkle[3], W. Bush[5], L-S. Wang[1], L. Farrer[4], J. Haines[5], R. Mayeux[2], M. Pericak-Vance[3], G. Schellenberg[1], Alzheimer's Disease Genetics Consortium; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Columbia Univ., New York, NY, [3]Univ. of Miami, Miami, FL, [4]Boston Univ., Boston, MA, [5]Case Western Reserve Univ., Cleveland, OH, [6]Cleveland Clinic Neurological Inst., Cleveland, OH

Abstract:

Increasing diversity in genomic studies is critical for identifying variants more prevalent in or specific to different ancestries to better characterize LOAD genetic architecture. We created and analyzed a multi-ancestry collection of GWAS in the Alzheimer's Disease Genetics Consortium (ADGC) to identify novel LOAD susceptibility loci and characterize shared and unique LOAD genomic risk profiles between ancestry groups. We included GWAS genotype and phenotype data on 38,774 non-Hispanic White (NHW), 7,454 African American (AA), 11,436 Hispanic (HI), and 3,277 East Asian (EAS) subjects, all imputed to the NHLBI TOPMed v5 reference panel. We performed a two-stage analysis: (1) single-variant association using score-based logistic regression for case-control and cohort studies and generalized linear mix-model for family-based datasets, followed by a within-ancestry fixed-effects meta-analysis using METAL; and (2) cross-ancestry meta-analysis of within-ancestry summary statistics using the random-effects model (RE2) in METASOFT. Covariates included onset/exam age, sex, and principal components for population substructure. In addition to *APOE* region associations, we observed eleven known loci with cross-ancestry genome-wide significant associations (P<5×10^{-8})
at/near *CR1*, *BIN1*, *TREM2*, *CD2AP*, *NYAP1*, *CLU*, *PTK2B*, *ECHDC3*, *MS4A6A*, *PICALM*, and *ABCA7*. Four novel loci were identified on chromosomes 2q24.3 (rs74439126 near *FIGN*/*GRB14*, $P=4.70\times10^{-8}$), 8q24.3 (rs34173062 in *SHARPIN*, $P=1.24\times10^{-9}$), 11p12 (rs12576934 near *LRRC4C*, $P=1.74\times10^{-8}$), and 12q24.13 (rs115185024 near *LHX5-AS1*, $P=3.68\times10^{-8}$). Highly heterogeneous association patterns were observed across ancestries: signals at 2q24.3 and 12q24.13 were driven by the HI ancestry group ($P=2.32\times10^{-9}$ and $P=7.66\times10^{-10}$, respectively), while for the 8q24.3 and 11p12 loci, suggestive associations were observed in one ancestry (NHW) ($P=1.38\times10^{-7}$ and P=1.81×10^{-6}, respectively) with nominal significance in the HI ancestry ($P=3.20\times10^{-4}$ and $P=0.04$). Follow-up analyses are in progress, including cross-ancestry fine-mapping, gene-based analyses, and eQTL analyses. These novel loci include strong biological candidates: *LRRC4C* (Leucine-Rich Repeat Containing 4C; MIM:608817) encodes a ligand influencing axon guidance and that regulates thalamocortical axon development/function, while *SHARPIN* (SHANK associated RH domain interactor; MIM: 611885) has been associated with dementia-related traits and is implicated in inflammatory pathways related to AD. Multi-ancestry studies with even larger sample sizes are necessary to further elucidate the genomic underpinnings of LOAD.

S22. Epigenomic associations with disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 128. Method for lung cancer detection using BAL fluidbased on differential methylation pattern analysis by MRE-seq

Authors:

N. Min[1], K. Lee[1], H. Kim[1], M-S. Jeong[1], T-W. Joo[1], S-Y. Yun[1], Y. Lim[1], D. Park[1], K. Lee[1], Y-H. Kim[1], H-J. Kwon[1], H. Kim[2], S. Kim[3], H. An[2], J. Shin[2], S. Kim[3], B. Shim[2], S. Lee[1]; [1]Eone Diagnomics Genome Ctr. Inc., Incheon, Korea, Republic of, [2]Div. of Med. Oncology, Dept. of Internal Med., St. Vincent's Hosp., Coll. of Med., The Catholic Univ. of Korea, Suwon, Korea, Republic of, [3]Div. of Pulmonary and Critical Care Med., Dept. of Internal Med., St. Vincent's Hosp., Coll. of Med., The Catholic Univ. of Korea, Suwon, Korea, Republic of

Abstract:

Lung cancer is one of the most common malignant tumors worldwide. With lung cancer having a poor prognosis, early diagnosis and treatment are essential for patient survival, as well as for precision medicine, efficacy monitoring, and prognosis prediction. However, there is no reliable method for detecting lung cancer in its early stage. Even though Bronchoalveolar lavage (BAL) fluid has emerged as a promising source of information for diagnosing lung cancer, its usage has been limited to diagnosing interstitial lung diseases and infectious diseases because there is no adequate method to detect lung cancer signal. Thus, we present an MRE-seq-based analysis that can distinguish lung cancer from benign disease through methylation pattern analysis of using BAL fluids and present its clinical usability.

For this study, 20 patients with lung cancer and 20 patients with benign disease were enrolled respectively. And BAL fluid was collected from each patient. DNA sample isolated from each BAL fluid was subjected to MRE-seq to analyze the methylation status of the DNA. A variety of machine learning algorithms such as XGBoost, AdaBoost and RGF were then used to determine the difference in methylation patterns between 20 lung cancer samples and 20 benign disease samples.

About 1000 markers were selected for each Leave-One-Out Cross-Validation using markers that showed at least twice the intensity in cancer samples compared to positive samples, and XGboost analysis was performed using these markers to obtain the most accurate results. AUC for lung cancer risk prediction model with BAL fluid is 0.96 which shows 95% sensitivity was achieved at 100% specificity.

Methylation profile analysis of DNA from BAL fluid has demonstrated excellent performance in distinguishing between lung cancer and benign diseases, and could be applied as a diagnostic tool.

S22. Epigenomic associations with disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 129. Multi-level epigenome profiling reveals a distinct role for DNA methylation in driving cellular differentiation

Authors:

E. Hodges[1,2], L. Guerin[1]; [1]Dept. of Biochemistry, Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Genetics Inst., Vanderbilt Univ., Nashville, TN

Abstract:

Cell differentiation involves coordinated regulation of the epigenome on multiple levels, from DNA methylation (DNAme) to chromatin accessibility (ChrAcc). Canonical models of gene regulation assume that ChrAcc and DNA demethylation are prerequisites for gene transcription. However, our recent work demonstrates that DNAme and chromatin dynamics are not as tightly linked as previously thought, challenging the causal relationship between DNAme and transcription.
Using ATAC-Me, a method developed by our lab to simultaneously profile DNAme and ChrAcc, we investigated the coordinated dynamics of DNAme and ChrAcc during a densely sampled time course of early neural progenitor cell differentiation. We show that ChrAcc responds quickly, and transiently in some genomic contexts, to induction of differentiation with ~38,000 regions displaying dynamic ChrAcc behavior. Many of these regions (the majority of which are lineage specific enhancers) display concordant changes, where decreases in DNAme accompany increases in ChrAcc. However, a substantial subset shows discordant temporal behaviors, where regions become hypomethylated despite opening and closing of chromatin. In contrast to ChrAcc and transcriptional changes that begin as early as 6 hours post-induction, the greatest loss in DNAme occurs several days later, primarily during a specific window of time coinciding with increased TET expression and peak 5-hydroxymethylation levels, confirming active removal of methylation at these sites. Using RNA-seq and transcription factor (TF) footprinting, we identified differentially active TFs and determined their binding states, finding specific TF binding patterns echo the relationship between ChrAcc and DNAme.
Overall, we show that a majority of lineage-specifying enhancers undergo periods of DNA demethylation that is temporally distinct from other regulatory events. Furthermore, hypomethylation of these regions persists long after TF binding and ChrAcc have dissipated, suggesting that long-lasting hypomethylation of certain enhancers is a historical record of previous activity. Our findings provide important context as to when DNAme exerts its regulatory function and suggests that DNA demethylation at a distinct subset of enhancers is a critical switch reinforcing phenotypic transitions during cell fate specification.

S22. Epigenomic associations with disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 130. Profiling pre-diagnosis plasma cell-free DNA methylomes up to seven years prior to clinical detection reveals early signatures of cancers

Authors:

N. Cheng[1,2], K. Skead[1,2], A. Singhawansa[3], S. Chen[3], T. Ouellette[1,2], H. He[3,4], D. Cescon[3,5], S. Bratman[3,4,6], D. DeCarvalho[3,4], D. Soave[7], P. Awadalla[1,2,5,8]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]Princess Margaret Cancer Ctr., Univ. Hlth.Network, Toronto, ON, Canada, [4]Dept. of Med. Biophysics, Univ. of Toronto, Toronto, ON, Canada, [5]Dept. of Med., Univ. of Toronto, Toronto, ON, Canada, [6]Dept. of Radiation Oncology, Univ. of Toronto, Toronto, ON, Canada, [7]Dept. of Mathematics, Wilfrid Laurier Univ., Waterloo, ON, Canada, [8]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada

Abstract:

Profiling cell-free DNA (cfDNA) patterns has emerged as a prominent non-invasive biomarker for early detection and subtyping of cancers. However, owing to difficulties in observing the early development of human malignancies as cancers are often detected once they become symptomatic, most cancer biomarker and evolution studies to date have primarily examined the genomics from solid tumour or liquid biopsies following a diagnosis. Utilizing cfDNA as a screening tool for early cancer detection requires profiling of blood plasma samples collected from asymptomatic individuals prior to the diagnosis of cancers to enable assessment of the earliest detectability and predictive performance of potential biomarkers. Here, we leverage the Canadian Partnership for Tomorrow's Health Project (CanPaTH), to profile blood plasma collected prior to the clinical detection of underlying cancers. Specifically, we use cell-free methylated DNA immunoprecipitation and high-throughput sequencing (cfMeDIP-seq), a highly sensitive assay for profiling cfDNA methylomes, to profile over 300 blood plasma samples collected up to seven years prior to the detection of a breast, prostate or pancreatic cancer, in addition to matched controls with no history of cancer free through follow-up. We identified differentially methylated signatures in pre-diagnosis cfDNA that discriminated cancer-free controls from pre-diagnosis cancer cases over five years before diagnosis, and demonstrated that these markers were reflective of methylation profiles from the originating cancer tissue. Further, predictive modelling reveals that cfDNA methylation markers in blood are predictive of pre-diagnosis breast cancer cases, achieving an AUROC of 0.852 among held-out pre-diagnosis test samples and 0.925 among an external test set of post-diagnosis breast cancer and control samples. Likewise, we demonstrate that cfDNA methylation markers predictive of early breast cancers were detectable among individuals that tested negative for a mammogram screening within one year of biologic collection and among individuals diagnosed before 50, preceding the recommended age for mammogram screening in Canada. Further, predictive models trained solely with pre-diagnosis cfDNA methylation samples were also generalizable and predictive of prostate and pancreatic cancer samples collected following diagnosis, achieving average test AUROCs of 0.95 and 0.96. In our current studies, we focus specifically on breast, prostate, and pancreatic cancer cases, and are extending this to further pan-cancer applications in subsequent investigations.

S22. Epigenomic associations with disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 131. MeOW: genome-wide identification of differentially methylated regions from Oxford Nanopore long-read sequencing data

Authors:

M. Galey[1], C. Paschal[2], A. Beck[3], E. Eichler[4], D. Miller[1]; [1]Univ. of Washington, Seattle, WA, [2]Seattle Children s, Seattle, WA, [3]Univ Washington & Seattle Children's Hosp., Seattle, WA, [4]Univ of Washington, Seattle, WA

Abstract:

Despite advances in both clinical and research genetic testing nearly 50% of individuals with a suspected genetic disorder remain undiagnosed after a complete evaluation. Many groups are interested in applying long-read sequencing (LRS) technology to these unsolved cases with the goal of identifying disease-causing variants either missed or difficult to detect with prior testing modalities. LRS is unique in that as a single data source it can be used to identify single nucleotide variants (SNVs), insertion or deletion variants (indels), structural variants (SVs), and differences in methylation. While many tools exist for the identification and analysis of SNVs, indels, and SVs, there are few, if any, able to perform genome-wide evaluation for differentially methylated regions (DMRs) that may contribute to disease.
Here we present Methylation Optimization Wizard (MeOW), a program to identify and prioritize DMRs genome-wide from Oxford Nanopore LRS data. MeOW can be run either with a file containing counts of per-nucleotide methylated CpG sites or with a bam file containing modified base tags in hts-specs format. MeOW runs quickly, taking approximately 20 minutes to complete analysis on a human genome at 40x coverage on a modestly powered computer. We have found that MeOW has 100% sensitivity for identifying DMRs in a cohort of individuals with known imprinting disorders such as Prader-Willi syndrome, Angelman syndrome, and Beckwith Wiedemann syndrome. Using MeOW on a cohort of individuals who have remained unsolved after comprehensive clinical testing revealed high-priority DMRs for additional evaluation. MeOW will simplify genome-wide analysis of challenging unsolved cases and permit identification of novel DMRs associated with human disease.

S22. Epigenomic associations with disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 132. Whole blood-based epigenome-wide association study of renal cell carcinoma reveals hypomethylation of a human-specific sequence in the second intron of *PCBD2*

Authors:

A. Shimizu[1], H. Ohmomo[2], S. Komaki[3], Y. Sutoh[4], T. Hachiya[5], K. Ono[1], E. Arai[6], H. Fujimoto[7], T. Yoshida[7], Y. Kanai[6], K. Asahi[1], M. Sasaki[1]; [1]Iwate Med. Univ., Shiwa-gun, Iwate, Japan, [2]Iwate Med. Univ., Shiwa-gun, Japan, [3]Iwate Tohoku Med. Megabank Organization, Yahaba, Iwate, Japan, [4]Iwate Med. Univ., Yahaba, Japan, [5]Iwate Med. Univ., Yahaba, Shiwa, Iwate, Japan, [6]Keio Univ., Tokyo, Japan, [7]Natl. Cancer Ctr. Res. Inst.., Tokyo, Japan

Abstract:

Background
Renal cell carcinoma (RCC) is the fourteenth most common cancer worldwide, accounting for approximately 4% of all cancers. More than 70% of RCC cases are clear cell RCC (ccRCC). To date, no reliable biomarkers for the detection of ccRCC have been identified. The aim of this study was to identify blood-based DNA methylation (DNAm) markers for the early detection and treatment of ccRCC.
Results
To identify ccRCC-associated DNAm markers, we performed targeted bisulfite sequencing (TB-seq) and an epigenome-wide association study (EWAS) using whole blood-derived DNA from 50 ccRCC patients and 50 healthy controls in the discovery phase. EWAS was performed using a linear regression model. The analysis was adjusted for age, sex, and the estimated cell-type composition. In the replication phase, the accuracy of the identified ccRCC-associated CpGs was verified in 48 independent ccRCC patients and 48 healthy controls. We identified six ccRCC-associated hypomethylated CpGs in the second intron of pterin-4 alpha-carbinolamine dehydratase 2 (*PCBD2*) in the discovery phase ($p < 1.75 \times 10^{-8}$), of which four were replicated in the replication phase ($p < 2.96 \times 10^{-8}$). The sum of the DNAm levels at the six CpGs was a valid indicator of ccRCC both in the discovery (area under the receiver operating characteristic curve [AUC-ROC] = 0.922) and in the replication phase (AUC-ROC = 0.871). Moreover, the results of cis-expression quantitative methylation analysis suggest that the DNAm levels of the ccRCC-associated CpGs affect the gene expression of transcription factor 7 (*TCF7*) and voltage-dependent anion-selective channel 1 (*VDAC1*), which are involved in cancer progression. Interestingly, homology searches revealed that this region is a sequence unique to the human genus, as it was not found to be present in chimpanzees or gorillas, and was presumed to be the result of a transposition from the mitochondrial genome into the nuclear genome.
Conclusions
In this study, we identified six ccRCC-associated CpGs in the human-specific second intron of *PCBD2* via EWAS using blood-derived DNA. We found that the DNAm level of these CpGs is a potential biomarker for early ccRCC detection; the value of this biomarker requires investigation in future studies.

S22. Epigenomic associations with disease

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 133. An epigenome-wide association study (EWAS) of allergic sensitization in children of diverse ancestry using a custom allergy & asthma array reveals an enrichment for differentially methylated high-value CpGs compared to the EPIC array

Authors:

E. Thompson[1], A. Morin[1], B. Helling[1], P. Faber[1], T. Gebretsadik[2], L. Bacharier[3], M. Kattan[4], G. T. O'Connor[5], K. Rivera-Spoljaric[6], R. A. Wood[7], M. C. Altman[8], T. Hartert[2], D. J. Jackson[9], J. Gern[9], C. McKennan[10], C. Ober[1]; [1]The Univ. of Chicago, Chicago, IL, [2]Vanderbilt Univ. Sch. of Med., Nashville, TN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Columbia Univ. Med. Ctr., New York, NY, [5]Boston Univ. Sch. of Med., Boston, MA, [6]Washington Univ., St Louis, MO, [7]Johns Hopkins Univ., Baltimore, MD, [8]Benaroya Res. Inst., Seattle, WA, [9]Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [10]Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

The epigenome plays a critical role in regulating gene expression in a context-specific manner, such as in the presence of environmental exposures or disease states. However, much of the DNA methylome remains unexplored, in part because our knowledge is largely limited to the CpGs on commercial arrays, which comprise <5% of CpGs in the human genome. We addressed this gap by using a custom Allergy&Asthma DNA methylation array (Morin et al. 2022; https://doi.org/10.1101/2022.05.19.22275204) to perform an EWAS of allergic sensitization (AS). The 37,863 CpGs on the Custom array were selected based on functional criteria and represent a set of high-value (likely functional) CpGs. DNA from nasal epithelial cells (NECs) was collected at age 11 from ethnically diverse children in The Urban Environment and Childhood Asthma (URECA) cohort and hybridized to both the Custom and EPIC arrays for comparison; data were processed in parallel using minfi. All EWAS were performed using a linear model in limma. Results from the Custom array were replicated in NECs from children in the Infant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure (INSPIRE) cohort (ages 5-7), a primarily non-Hispanic white population. The Custom array was enriched for differentially methylated CpGs (DMCs) associated with AS compared to the EPIC (q-value<0.05; 0.50% vs. 0.23%; Fisher's Exact Test (FET) $p<2.2x10^{-16}$). Among the AS DMCs on the Custom array, 13% were also DMCs in INSPIRE (q-value<0.05), which was a highly significant enrichment ($p<2.2x10^{-16}$); all had concordant directions of effect. The effect sizes of the CpGs that were DMCs in either URECA or INSPIRE were also highly correlated (r=0.61; $p<2.2x10^{-16}$). We used gene expression data collected in the same NECs from URECA children and defined two sets of genes from among the 15,551 detected as expressed: the nearest gene to each CpG and the promoter capture Hi-C (pcHi-C)-defined target gene. For each CpG-gene pair, we tested for correlation between DNA methylation and expression levels using a linear model to identify expression quantitative methylation (eQTM) CpGs (q<0.05). We observed an over-representation of eQTMs among Custom DMCs compared to EPIC DMCs for both nearest (35% vs. 20%, FET p=0.0019) and pcHi-C target genes (22% vs. 15%, FET p=0.0082). The high reproducibility of EWAS results in two cohorts demonstrates that CpGs on the Custom array identify AS DMCs that are robust to ancestry, ascertainment strategy, age at sampling, and geography. Overall, our findings show that CpGs on the Custom array contribute to allergic phenotypes and that CpGs in currently invisible portions of the epigenome are relevant to human health.

S23. Let's talk about sex and its contributions to disease risk

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 136. Testing for interaction using over 1 million individuals and 20 diseases shows considerable differences in polygenic risk score estimates across age groups and sexes

Authors:

B. Jermy[1], B. Wolford[2], K. Läll[3], T. Hartonen[1], Z. Yang[1], R. Monti[4], J. Wanner[4,1], Y. Wang[5], K. Zguro[6], O. Youssef[7], FinnGen, S. Furini[6], A. Martin[5], H. Heyne[4,1,8], K. Hveem[2], R. Mägi[3], A. Palotie[1], N. Mars[1,5], A. Ganna[1,5], S. Ripatti[1,5,9]; [1]Inst. for Molecular Med. Finland, FIMM, Helsinki, Finland, [2]NTNU, Trondheim, Norway, [3]Estonian Genome Ctr., Inst. of Genomics, Tartu, Estonia, [4]Hasso Plattner Inst., Digital Hlth.Ctr., Potsdam, Germany, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Univ. of Siena, Siena, Italy, [7]Univ. of Helsinki, Helsinki, Finland, [8]Hasso Plattner Inst., Mount Sinai Sch. of Med., New York, NY, [9]Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland

Abstract:

Absolute lifetime risks for polygenic risk score (PRS) strata are important for clinical decision making. Prior research typically assumes PRS effects are constant across key demographic groups such as age and sex although age and sex are key risk factors for many diseases. We applied our novel method for estimating age and sex specific lifetime risks (cumulative risk by age 80) across PRS strata in over 1 million participants from 4 biobanks (UK Biobank (UKB), FinnGen, Estonian Biobank (EstBB), and Trondelag Health Study (HUNT)).

Twenty diseases were selected to have a large impact to the Global Burden of Disease (GBD), moderate heritability and well-powered GWAS. Age and sex stratified effects of PRS on disease were assessed using Cox proportional hazards and sex specific baseline risks were calculated using publicly available data from the GBD.

PRS had a considerably increased HR in younger age groups for 7 diseases with evidence of replication across biobanks. For example, atrial fibrillation was more strongly associated in younger age groups for all three biobanks tested (HRs Youngest vs Oldest Quartile in FinnGen = 5.42 (95% CI: 4.83 - 6.09) vs 2.79 (2.26 - 3.44); EstBB = 5.09 (4.33 - 5.97) vs 1.14 (0.46 - 2.8); HUNT = 9.67 (7.3 - 12.8) vs 2.51 (1.32 - 4.78)). Notably, this finding appears to be disease specific with numerous diseases showing consistent effects across age, including breast cancer.

For diseases demonstrating a stronger effect with age, age-specific HRs increased cumulative risk at younger ages relative to assuming a constant effect across age. For type 2 diabetes, each 10% increment in cumulative risk was attained in the prior 5-year age bin for both Estonia and Finland relative to assuming a constant effect.

Four diseases (knee osteoarthritis, coronary heart disease, gout, and rheumatoid arthritis) also repeatedly exhibited sex differences in all biobanks tested. For males and females in the top 1% of polygenic risk respectively, lifetime risks for gout were 27.6% and 9.2% in Finland, 30.5% and 6.5% in Norway, and 37.6% and 9.9% in the UK. No significant differences in PRS hazard ratios (HR) were detected by sex; indicating variation in baseline risk by sex is the key differentiator.

Accurate lifetime risk estimation optimizes clinical implementation. In this study, we show lifetime risk can vary substantially by sex, even if the relative risk of the PRS is consistent between sexes. Further, we show for many common diseases, age-stratified analyses increase the risk at earlier ages. Such differences may lead to differential screening for individuals with high polygenic risk, however, as we show, the degree of variation is disease specific.

S23. Let's talk about sex and its contributions to disease risk

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 137. GWAS meta-analysis of WHRadjBMI identifies differences across sexes and populations

Authors:

E. Wilson[1], S. Vedantam[2,3], Z. Fairhurst-Hunter[4], E. Marouli[5], J. Arias[6], G. Chittoor[7], S. Berndt[6], T. Winkler[8], K. Young[9], M. Graff[9], C-T. Liu[10], R. Walters[4,11], C. Lindgren[12,13], K. Mohlke[1], A. Justice[7], Genetic Investigation of ANthropometric Traits (GIANT) Consortium; [1]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Dept. of Endocrinology, Boston Children's Hosp., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Dept. of Population Hlth., Big Data Inst. Building, Univ. of Oxford, Old Road Campus, Roosevelt Drive, Oxford, United Kingdom, [5]William Harvey Res. Inst., Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [6]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, [7]Dept. of Population Hlth.Sci., Geisinger Hlth.System, Danville, PA, [8]Dept. of Genetic Epidemiology, Univ. of Regensburg, Regensburg, Germany, [9]Dept. of Epidemiology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [10]Dept. of Biostatistics, Boston Univ., Boston, MA, [11]MRC Population Hlth.Res. Unit, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [12]Wellcome Trust Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [13]Oxford Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Univ. of Oxford, Oxford, United Kingdom

Abstract:

Genetic associations for waist-to-hip ratio adjusted for body mass index (WHRadjBMI) can differ between sexes and populations, but the mechanisms underlying these differences remain unknown. We performed WHRadjBMI GWAS meta-analyses in >1.1 million individuals (56% female) from 201 studies that contributed GWAS results to the GIANT consortium. Initially, we identified conditionally distinct signals ($p<5e-9$) in 18 meta-analyses: females only, males only, and sex-combined; in European (947K), East Asian (136K), South Asian (41K), Hispanic/Latino (33K), African (28K), and all populations. We compared autosomal signals across the sex-combined, female, and male analyses within each population using an LD threshold of $r^2>0.8$ in ancestry-matched reference panels; the population-specific analyses identified 954 signals in European, 26 in East Asian, 0 in South Asian, 1 in Hispanic/Latino, and 2 in African populations and 1,026 signals in the initial multi-population analysis. When we combined signals across all 18 analyses, we identified a total of 1,280 signals, including 237, 12 and 1 signal identified only in European, East Asian, and African populations, respectively, highlighting the value of performing both population-stratified and multi-population analyses. Preliminary multi-population fine-mapping identified 6 signals with differences in LD structure between populations that may help to identify causal variants driving the signals. For example, a signal near *ADAMTS9* is 20% narrower and contains fewer proxy variants in East Asian populations than European populations. When considering heterogeneity between the sexes, effect sizes at 487 (47%) of the multi-population association signals were heterogeneous (FDR 0.05); 84% of which showed stronger association in females than males. Population-specific analyses showed similarly that 49% of signals identified in European individuals and 62% of signals identified in East Asian individuals showed significant sex heterogeneity. One example sex-heterogeneous WHRadjBMI signal near *FAM13A* showed a very strong multi-population association in females ($p=6e-27$) but not males ($p=0.02$); in GTEx subcutaneous adipose tissue, *FAM13A* showed sex-biased gene expression, higher in females (median 20.3 TPM) than males (19.0 TPM), suggesting a potential basis for sex differences at the locus. Overall, this GWAS meta-analyses greatly expands the number of identified WHRadjBMI signals, providing opportunities to identify the underlying genes and mechanisms, and further demonstrates the need to perform genetic analyses in diverse populations, including those stratified by sex.

S23. Let's talk about sex and its contributions to disease risk

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 138. Phenome-wide sex differences in environmental effects in the National Health and Nutrition Examination Survey (NHANES) 1999-2006

Authors:

T. Gonzalez Zarzar, N. E. Palmiero, M. A. Hall; The Pennsylvania State Univ., State College, PA

Abstract:

Sex and gender differences are ubiquitous in health and disease, including differences in exposure effects on endophenotypes and health outcomes. A phenomic environment-wide study (PheEWAS) can generate novel hypotheses on sex-specific environmental effects. We used a curated set of 1,191 phenotypes and environmental variables and 41,474 participants across four survey periods (1999-2006) of the National Health and Nutrition Examination Survey (NHANES). The data sets were divided into discovery and replication cohorts, and each was further divided into female and male cohorts. A standard quality control procedure was applied to each cohort and completed with a set of 10 covariates, 58 phenotypes, and 272 exposures. A generalized linear regression was fitted for each phenotype and exposure, controlling for ten covariates and adjusting for survey weights. Sex differences were identified by estimating the difference in effect sizes between females and males and were categorized into pure (significant effect in a single sex), quantitative (differences in effect sizes), and qualitative (differences in effect directions) by evaluating the nominal significance and direction of effect of each exposure-phenotype association. 119 of the 15,466 regression tests were considered significantly different between sexes, of which 49 were classified as pure, 45 as quantitative, and 25 as qualitative sex differences. For example, the volatile compounds blood ethylbenzene ($\beta f = 0.096$, $\beta m = -0.020$, P=1.54E-5) and blood toluene ($\beta f = 0.115$, $\beta m = -0.015$, P = 3.91E-6) were significantly associated with red cell count in women only. On the other hand, cadmium exposure had greater effect sizes in c-reactive protein ($\beta f = 0.045$, $\beta m = 0.144$, P = 6.01E-9), and monocyte number ($\beta f = 0.092$, $\beta m = 0.166$, P = 7.48E-6) in men compared to women. Smoking-related exposures were among the environmental exposures showing the greatest number of sex differences. Women showed greater positive effect sizes of smoking exposure on homocysteine levels ($\beta f = 0.154$, $\beta m = 0.038$, P = 2.71E-6), and hemoglobin ($\beta f = 0.429$, $\beta m = 0.236$, P = 1.25E-6), which can be a mechanism that explains the greater risk of cardiovascular diseases in women who smoke. Conversely, men showed greater negative effect sizes of smoking exposure on albumin ($\beta f = 0.005$, $\beta m = -0.089$, P = 2.36E-13) and bilirubin levels ($\beta f = -0.077$, $\beta m = -0.218$, P = 2.42E-6), which might highlight a potential heightened inflammatory response to smoking exposure in men. Taking environmental exposures and sex differences as fundamental pieces in the development of complex diseases is relevant for accurate and precise methodologies that aim to predict them.

S23. Let's talk about sex and its contributions to disease risk

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 139. Regulatory network approaches reveal sex differences in cancer gene regulation

Authors:

C. Lopes-Ramos[1,2,3], J. Quackenbush[3], D. DeMeo[2,1]; [1]Harvard Med. Sch., Boston, MA, [2]Brigham & Women's Hosp, Boston, MA, [3]Harvard TH Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Despite established sex differences in epidemiology and clinical presentation of many diseases, including cancer, most preventive and therapeutic strategies do not account for sex differences. This is in large part because the molecular features that drive sex differences are poorly understood, and there are still conceptual and methodological gaps to incorporating sex into research and clinical practice. Our goal was to investigate transcriptional regulation of cancer genes in males and females across normal tissues to better understand sex differences in cancer risk and incidence rates. **Methods:** We developed a sex-specific network approach to integrate genomic data and model gene regulation in males and females. We used gene expression data from 29 healthy tissues in the GTEx dataset to infer subject-specific gene regulatory networks, which capture the interactions between transcription factors and their target genes. To uncover sex-biased regulatory processes present in normal tissues, we used network centrality measures (degree and closeness) to compare the networks between males and females, focusing on interactions between 486 cancer genes (as defined by the COSMIC Cancer Gene Census v92). Finally, we used cascade model to identify transcription factor master regulators driving the sex-biased regulatory processes. **Results:** We observed that cancer genes have greater indegree and closeness than non-cancer genes (t-test, $p<0.05$), indicating that they have more direct interactions with transcription factors and shortest network distance to all other genes in the network, respectively. We tested for sex differences in the cancer gene regulatory network, and found that oncogenes and tumor suppressor genes are differentially targeted by transcription factors in males and females, including tumor suppressor genes that escape X chromosome inactivation in females (t-test, $p<0.05$). Finally, we discovered an average of 20 sex-biased transcription factor master regulators across all 29 tissues (z-score sex difference > 2). Using Gene Set Enrichment Analysis, we found that sex-biased master regulators target cancer-related pathways, including WNT, NOTCH, and p53 (FDR<0.05). **Conclusion:** Our findings showed that normal tissues have sex-biased regulation of genes implicated in tumorigenesis, which might help explain the molecular basis of sex differences observed in cancer risk and incidence rates. Our sex-specific gene regulatory network approach can be extended to cancer datasets to understand how sex influences cancer progression and therapeutic responses and may help advance sex-aware precision oncology.

S23. Let's talk about sex and its contributions to disease risk

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 140. *ADGRG6* is involved in gender-specific fat distribution

Authors:

H. Nguyen[1], A. Ushiki[1], R. Sheng[1], H. Choquet[2], T. Hoffman[1], R. S. Gray[3], N. Ahituv[1]; [1]UCSF, San Francisco, CA, [2]Kaiser Permanente, Oakland, CA, [3]UT Austin, Austin, TX

Abstract:

Fat distribution differences between men and women are a major risk factor for metabolic disease, but their genetic etiology remains largely unknown. A recent genome-wide association study (GWAS) of body fat distribution identified multiple loci that are gender-specific, including a locus in an intron of the gene *ADGRG6*, a G-protein coupled receptor. However, the mechanisms underlying these dimorphisms remain largely unknown. Here, we conduct a gender-specific genetic association analyses in the UK Biobank cohort and find that an intronic SNP in *ADGRG6*, rs9403383, is genome-wide significantly associated with trunk fat mass in women ($p= 5.03 \times 10^{-13}$), but not in men ($p= 0.06$). We show that the female trunk fat associated variant of rs9403383 significantly reduces enhancer activity in preadipocytes. Deletion of *ADGRG6*, the enhancer encompassing rs9403383 or the associated SNP all impair human adipogenesis. Conditionally knocking out *Adgrg6* in mouse adipocytes or deleting its gender-associated adipocyte enhancer leads to female-like fat distribution in males, which are protected against high-fat-diet-induced obesity and have improved insulin response. To showcase its therapeutic potential, we further demonstrate that CRISPRi targeting of the *Adgrg6* promoter or adipocyte enhancer prevents high-fat-diet-induced obesity. Combined, our results associate *ADGRG6* as a gender fat distribution gene and highlight its potential as a therapeutic target for metabolic disease.

S23. Let's talk about sex and its contributions to disease risk

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 141. A proteome-wide investigation of sex differences in genetic control of brain protein abundance in 1100 human brain samples

Authors:

T. Wingo[1], Y. Liu[1], D. Duong[1], D. Cutler[1], M. Epstein[1], D. Bennett[2], A. Levey[1], N. Seyfried[1], A. Wingo[1]; [1]Emory Univ., Atlanta, GA, [2]Rush Univ., Chicago, IL

Abstract:

Background Why women are disproportionally affected by certain brain illnesses (for instance depression or Alzheimer's disease) while men by others (autism or Parkinson's disease) is not well understood. We hypothesize that sex differences in brain protein expression may contribute to these phenomena. Here, we investigated whether there are sex differences in brain protein abundance and in the genetic control of brain protein expression. **Methods** We used deeply profiled brain proteomes of 1100 donors of European ancestry and their genome-wide genotyping. The human brain proteomes were mostly from the dorsolateral prefrontal cortex (dPFC). We performed quality control and normalization of the proteomic data and surrogate variable analysis (SVA). Then, we regressed out effects of age, batch, post-mortem interval, clinical diagnosis, and SVAs from the proteomic profile before examining effect of sex on brain protein expression and on genetic control of brain protein expression using linear regression. **Results:** Of the 10,198 proteins detected in human brain, 12% (n=1239 proteins) exhibited sex-difference in abundance level. About half of these proteins had higher abundance and half had lower abundance in men compared to women after adjusting for effects of age and potentially hidden confounding factors. Moreover, 3.5% of these sex-difference proteins were encoded by the X chromosome. We examined potential sex difference in the genetic control of brain protein expression in the dPFC region. We found genome-wide significant SNP-by-sex interactions on protein abundance in 1171 SNP-protein pairs (at FDR $p<0.05$, N=793, specifically 150 unique proteins and 166 unique SNPs). As expected, approximately 48% of these SNPs were associated higher expression and 52% with lower expression of brain proteins in men compared to women after considering effects of age and potentially hidden confounding factors. Next, we examined these 166 SNPs in each sex separately and found that 35% were pQTLs in men only, 14% were pQTLs in women only, 7% were pQTLs in both sexes with same directions of association, and 41% were pQTLs in both sexes but with opposite directions of association (at $p<0.05$). **Discussion:** Our study provides here novel evidence of differences in protein abundance and genetic control of protein abundance in the human brain. This large, comprehensive sex-specific protein and pQTL data resource can facilitate studies investigating genetic and molecular underpinnings of sex differences in psychiatric and neurologic diseases and contribute to the identification of sex-aware treatment targets for these debilitating brain illnesses.

S24. Massively parallel variant characterization

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 144. Deciphering sequence effects of splicing in health and disease with massively parallel reporter assays

Authors:

A. M. Yu[1], S. Koplik[1], C. M. Roco[2], Y. Zhang[1], N. Bogard[3], A. K. Sabo[1], A. B. Rosenberg[2], G. Seelig[1]; [1]Univ. of Washington, Seattle, WA, [2]Parse BioSci.s, Seattle, WA, [3]Modulus Therapeutics, Seattle, WA

Abstract:

Alternative splicing (AS) is a key mechanism of eukaryotic gene expression diversity and can lead to many protein isoforms from single gene sequences. Dysregulation of AS is implicated in many human diseases such as cancer and spinal muscular atrophy. AS is regulated by a *cis*-regulatory sequence code that is interpreted by *trans*-acting RNA binding proteins with their own cell type-specific expression. Disentangling the individual *cis*- and *trans*-regulatory element effects in AS in the endogenous context has proven challenging due to the limited ability to measure sequence effects that are not evolutionarily selected. However, massively parallel reporter assays (MPRAs) have proven to be a useful tool for simultaneously assaying molecular phenotypes of thousands to millions of sequence combinations in multiple cell types. Here we focus on exon skipping (ES), the most common form of AS, and have developed a MPRA containing over 92,000 sequences derived from over a thousand of human genes. We have successfully transfected it into several types of mammalian cells. The MPRA encompasses short human exons and their flanking introns as well as variants thereof. ClinVar, Geuvadis and Exome Aggregation Consortium (ExAC) databases were used to identify disease-associated and common human variants. Site saturation mutagenesis (SSM) of several of these native sequences were also included to expand assayed sequence diversity beyond those observed in human populations. The MPRA spans 2,152 exons from 1,748 genes and contains 68 variants from Geuvadis, 6,736 from ExAC, and 83,335 sequences from disease-associated genes. The 83,335 sequences include ClinVar variants and many additional variants. We have fully characterized the MPRA library in three cell lines. HEK293 replicates show high reproducibility with the majority of reporters achieving high coverage in bulk RNA-seq. Comparing our HEK293 data to our datasets from HeLa and K562, we see a spread of ES phenotypes as well as some differential splicing patterns between cell lines from the same sequences. These datasets will be used in machine learning models to build generalizable and predictive understanding of ES in diverse human cell types. MPRAs provide rich datasets for machine learning algorithms to learn sequence variation effects that are not necessarily limited to observed population variants. These resulting machine learning algorithms provide more generalizable and predictive modeling of sequence to function relationships. We will further explore the ability of employing these datasets and models to engineer RNA splicing behaviors for translational understanding of human disease.

S24. Massively parallel variant characterization

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 145. Improving estimation of variant functional effects from high-throughput assays

Authors:

T. Yu[1], J. Fife[1], R. Sherwood[2], C. Cassa[3]; [1]Brigham and Women's Hosp., Boston, MA, [2]Brigham and Women's Hosp., Harvard Med Sch., Boston, MA, [3]Harvard Med. Sch., Boston, MA

Abstract:

Scalable CRISPR-Cas9 techniques such as deep mutational scanning and base editing now enable cost effective functional assessment of coding variants. However, experimental conditions and editing efficiency of these assays are far from uniform, which can result in confounding noise when predicting functional effects at the individual variant level. Given that functional screens often consist of multiple measurements of the same amino acid substitution across a protein, as well as repeated measurements at the same codon, we established a computational pipeline to use this information collectively to improve estimation of functional effects. Drawing from a large corpus of deep mutational scanning data (DMS), we first estimate the mean functional effects per codon within each gene (James-Stein estimator). We then normalize functional effects per assay and per position to build an amino acid substitution matrix (FUNSUM), and then use it to make estimates for individual allelic variants using both the positional context and specific substitution. To demonstrate the pipeline's capability to improve estimation for various functional assays, we estimated functional scores for coding variants in a BRCA1 base editing assay, a TP53 DMS assay and a muti-gene base editing assay for DNA Damage Response (DDR) genes. We first evaluated the improvement of our estimation pipeline on clinically significant variants (Clinvar), using separation in the score distributions between variants previously classified as pathogenic (P/LP) or benign (B/LB). In all three assays, we observed greater significance in separation between P/LP and B/LB variant groups using our estimated functional scores (KS, BRCA1 p=2.24e-51, TP53 p=1.78e-9, DDR genes p=1.63e-4) than with original functional scores (KS, BRCA1 p=2.42e-4, TP53 p=5.38e-8, DDR genes p=0.01). Similarly, we assessed estimation improvement for BRCA1 and TP53 assays using carriers of variants with or without associated phenotypes (UK Biobank). For the BRCA1 assay, the pipeline increased the number of carriers for whom predictions can be made from functional data (528 to 1400), and better separates patients who develop early breast cancer from those without breast cancer (KS, p=0.0238). For the TP53 assay, estimated functional scores were also better at separating patients who develop Li-Fraumeni syndrome (LFS) cancers from those without LFS cancer (p=2.96e-6). In conclusion, this approach promises to improve the quality and broaden the utility of data generated from CRISPR-Cas9 screening assays.

S24. Massively parallel variant characterization

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 146. Massively parallel functional assessment of missense variants in cardiac arrhythmia genes *KCNQ1* and *KCNE1* elaborates structure-function relationships

Authors:

A. Muhammad[1], T. Yang[2], M. Calandranis[2], M. Blair[2], L. Hall[2], K. Matreyek[3], D. Fowler[4], D. Roden[2], A. Glazer[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [4]Univ of Washington, Seattle, WA

Abstract:

Introduction: Loss- or gain-of-function variants in cardiac K^+ channel genes *KCNE1* and *KCNQ1* can cause fatal arrhythmias. Most observed variants are of uncertain clinical significance. **Objective:** To conduct massively parallel assays on comprehensive KCNE1 and KCNQ1 variant libraries to measure cell-surface trafficking (KCNE1, KCNQ1) and K^+ efflux-dependent cell death (KCNE1). **Methods:** We created comprehensive variant libraries of extracellularly tagged cDNA of 129-residue KCNE1 and 676-residue KCNQ1. From each library, single alleles were integrated per cell in a landing pad HEK293 line, used unchanged (KCNQ1 trafficking) or engineered to stably express KCNQ1-WT (KCNE1 trafficking) or gain-of-function KCNQ1-S140G (KCNE1-dependent K^+ conductance). For trafficking assays, cells were stained with a fluorescently tagged antibody and FACS-sorted into 4 equal groups. We scored variants using a weighted average of frequencies across 4 groups. For KCNE1 K^+ conductance assay, we hypothesized that the S140G variant would bias channels with functional KCNE1 to the open state, allowing constant K^+ efflux to reduce cell survival. Variants were scores using a $\log_2$-transformed ratio of frequencies in cells growing for 12 days to those in the original library. **Results:** Nonsense KCNE1 variants before residue 58 (early) were trafficking-deficient but after residue 57 (late) had WT-like trafficking. However, conductance was equally reduced in both early and most late nonsense variants. We hypothesize that late nonsense variants likely disrupt KCNQ1/KCNE1 interaction, rather than surface trafficking, to alter channel function. Furthermore, nonsense variants after residue 110 had WT-like K^+ conductance and are likely dispensable for channel function. For 1,886 KCNE1 missense variants, trafficking scores were normally distributed, but conductance scores showed a bimodal distribution; the two modes correspond to the means of synonymous and nonsense variants. In both assays, hydrophilic variation in the transmembrane domain was not tolerated. The KCNQ1 library represents 85% of 12,844 possible variants. Cell surface staining of the library of a 200-residue region spanning the transmembrane domain showed a bimodal distribution; the modes correspond to the means of WT and trafficking-null variants. Sequencing of these sorted pools is in progress. Future work will assess the rest of the protein and develop a conductance assay. **Conclusion:** Our work represents the first deep mutational scan of an arrhythmia-associated ion channel complex, encoded by *KCNQ1* and *KCNE1*. These data provide structural insights and may facilitate clinical variant interpretation.

S24. Massively parallel variant characterization

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 147. Saturation genome editing reveals 10% of missense SNV alleles in functional domains of PALB2 as functionally abnormal

Authors:

M. Dawood[1,2,3,4], I. Hill[5,6], S. Casadei[5,6], A. Mehrotra[5,6], N. Smith[5,6], K. de Leon[5,6], R. Daza[5,6], R. Garge[5,6], A. McEwen[5,6], S. Fayer[5,6], L. Muffley[5,6], J. Posey[1], J. Lupski[1,2], R. Gibbs[1,2], J. Shendure[5,6,7], D. Fowler[5,6], L. Starita[5,6]; [1]Baylor Coll. of Med., Houston, TX, [2]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [3]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [4]Med. Scientist Training Program, Baylor Coll. of Med., Houston, TX, [5]Univ. of Washington, Seattle, WA, [6]Brotman Baty Inst., Univ. of Washington, Seattle, WA, [7]Howard Hughes Med. Inst., Seattle, WA

Abstract:

Partner and Localizer of BRCA2 (*PALB2*) functions in homology-directed repair of double stranded DNA breaks, promotes nuclear localization and stability of BRCA2, and enables cell cycle checkpoint functions. Pathogenic variants in *PALB2* have been implicated in breast, ovarian, and pancreatic cancer as well as Fanconi anemia, a bi-allelic rare disease trait (MIM: 610355). In 2021, *PALB2* was elevated to the ACMG's 73 medically-actionable genes in which secondary findings should be reported to patient providers. However, with 99% of >1875 *PALB2* missense single nucleotide variant (SNV) alleles classified as Variants of Uncertain Significance (VUS) in ClinVar, the majority of variants in *PALB2* still remain uninterpretable and thus not practically implementable for guiding patient management. We performed Saturation Genome Editing (SGE) to investigate the effect of all possible 10,683 SNVs in *PALB2* on cell survival. Within the N and C terminal functional domains required for binding RAD51, BRCA1, and BRCA2, 12% of missense SNV alleles result in decreased *PALB2* function that compromises cellular viability. The SGE-derived functional scores are 100% consistent with published orthogonal functional assay data and recapitulate known genomic phenomena such as nonsense SNV alleles escaping from nonsense-mediated decay and stop-loss SNV alleles escaping from non-stop decay. Further, SGE consistently and accurately distinguishes known pathogenic from benign SNV alleles in ClinVar with a known clinical significance for *PALB2* (currently, nonsense vs. synonymous SNV alleles). Additionally, SGE functional scores are highly correlated with *in silico* CADD predictions of pathogenicity and rare allele frequencies per gnomAD. Also, due to advances in SGE technology, we are able to delineate a subpopulation of missense SNV alleles that demonstrates loss of function at a slower rate supporting use of this method to identify putative hypomorphic alleles. In total, we demonstrate SGE functional scores are poised to have a major impact on clinical variant interpretation for *PALB2*.

S24. Massively parallel variant characterization

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 148. Massively parallel screen for rare 3' UTR variants regulating mRNA abundance

Authors:

T. Fu, K. Amoah, T. W. Chan, J-H. Bahn, J-H. Lee, S. Terrazas, R. Cheung, S. Kosuri, X. Xiao; Univ. of California, Los Angeles, Los Angeles, CA

Abstract:

Genome-wide association studies have revealed that most disease-associated genetic variants are in non-coding regions of the genome. It remains a major challenge to understand the functional impact of these non-coding variants, especially rare variants which constitute the majority of non-coding variants with unknown function. To fill in this gap, we developed a highly reproducible massively parallel screen for rare 3' UTR variants that may affect mRNA abundance. With this platform, we tested 14,490/14,494 rare variants and identified 3,066/3,944 functional candidates that regulate mRNA abundance in HEK293/HeLa cells. By overlapping the functional variants with the binding sites of miRNAs and RBPs as well as predicting novel binding motifs, we annotated the functional variants with potential mechanisms in regulating 3' UTR activity. We showed that these functional variants are disease-relevant and enriched in genes associated with cancer. Further, we validated the functional variants as gene expression outliers in TCGA cancer patients. Through prime editing in HEK293T cells, we generated genome-edited homozygous clones with the reference or alternative alleles for two functional variants located in *MFN2* and *FOSL2*. We observed that the functional variants in the 3' UTR of *MFN2* and *FOSL2* increased mRNA stability for both genes, consistent with the increased mRNA abundance of the variant alleles in the MPRA. Importantly, both variants led to significant alterations in cell proliferation (reduced cell proliferation by the *MFN2* variant, and increased cell proliferation by the *FOSL2* variant). Our findings illustrate the usability of our method to understand disease-relevant variants, identify causal variants and uncover the associated functional mechanisms of post-transcriptional regulation.

S24. Massively parallel variant characterization

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 149. Quantifying the functional effects of 234,448 likely causal regulatory variants underlying complex human traits

Authors:

L. Siraj[1], J. Ulirsch[2], H. Dewey[3], S. Kales[3], M. Kanai[2], D. Berenzy[3], K. Mouri[3], E. S. Lander[2], P. C. Sabeti[1], S. K. Reilly[4], H. Finucane[5,2], R. Tewhey[3]; [1]Harvard, Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]The Jackson Lab., Bar Harbor, ME, [4]Dept. of Genetics, Yale Sch. of Med., New Haven, CT, [5]Massachusetts Gen. Hosp., Boston, MA

Abstract:

Up to 90% of causal SNPs from genome-wide association studies (GWASs) fall in accessible chromatin in relevant cell types, pointing to the importance of regulatory element function in human health and disease. Yet, our understanding of how single nucleotide changes affect the activity of regulatory elements is incomplete. We aimed to uncover genomic mechanisms governing regulatory element function by testing the activity of 234,448 fine-mapped variants and more than 50k matched controls in a Massively Parallel Reporter Assay (MPRA), the largest robust assessment of variant effects by a high-throughput reporter assay to date.

In order to define a pool of likely causal variants affecting gene regulation, we used the results of our recent large-scale fine-mapping in GTEx v8, UK Biobank, and Biobank Japan [Kanai et al. *medRxiv* 2021]. In total, we included 17,407 variants with a high posterior inclusion probability (PIP) of being causal and 91,412 95% credible sets. We selected control variants to match test variant annotations or genomic locations. We assayed variants across 6 diverse cell lines (SK-N-SH, HepG2, K562, GM12878, A549, and HCT116). Within the eQTL set, we observed 55,402 elements that regulate gene expression in the MPRA. Within these active elements, we detected 18,288 variants that exhibit significant allele specific activity in the MPRA (expression-modulating variants; emVars).

We validated MPRA's ability to identify causal regulatory variants using high-PIP eQTLs as a gold standard. Non-coding high-PIP variants are 2.6 times more likely to be emVars (18%) as compared to low-PIP variants (7.4%) or controls (8.1%). Overall, variants that were emVars and within accessible chromatin (80% precision, 13% recall) outperformed chromatin accessibility alone (66% precision, 46% recall), element activity only (60% precision, 41% recall), and allele specific activity alone (68% precision, 18% recall) for identifying high-PIP eQTLs.

We next sought insight into the mechanisms of emVars. High-PIP emVars are most strongly enriched at promoters and multi-tissue enhancers, consistent with our finding that 80% of allelic effects are shared across more than one cell line. These variants are more significantly occupied by 279 diverse transcription factors (TFs) than high-PIP variants without allelic effects, but only 15% of high-PIP emVars disrupt a known binding motif for an occupied TF. We found no evidence of widespread multiple causal variants in a single 95% CS after controlling for MPRA background rate. In conclusion, MPRA provides a powerful approach to identify and dissect non-coding regulatory variants underlying human health.

S25. Novel statistical genetics methods for complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 152. A unified method for estimating direct genetic effects and performing genome-wide association studies

Authors:

J. Guan[1], S. Nehzati[1], H. Jayashankar[2], P. Turley[3,4], D. Cesarini[2,5], D. Benjamin[1,2,6], A. Young[1]; [1]UCLA Anderson Sch. of Management, Los Angeles, CA, [2]Natl. Bureau of Economic Res., Cambridge, MA, [3]Dept. of Economics, Univ. of Southern California, Los Angeles, CA, [4]Ctr. for Economic and Social Res., Univ. of Southern California, Los Angeles, CA, [5]Dept. of Economics, New York Univ., New York, NY, [6]Human Genetics Dept., UCLA David Geffen Sch. of Med., Los Angeles, CA

Abstract:

A goal of genome-wide association studies (GWAS) is to estimate the causal effects of alleles carried by an individual on that individual ("direct genetic effects"). However, GWAS actually estimates "population effects," which include indirect genetic effects (e.g. effects of alleles in parents on offspring through the environment) and bias from confounding due to population stratification and assortative mating. Obtaining unbiased estimates of direct genetic effects is important for removing bias from Mendelian Randomization analyses, studies of selection, estimates of heritability and genetic correlations, and polygenic prediction. Genetic variation within a family is random, so one can obtain unbiased estimates of direct genetic effects from individuals with both parents genotyped. However, parental genotypes are often missing. By imputing the genotypes of missing parent(s) from observed parent and offspring genotypes, as in the software package *snipar*, one can obtain estimates of direct genetic effects from any individual with at least one genotyped sibling or parent. However, this is typically only a small fraction of the genotyped individuals in large-scale biobanks. We show that the sample of 'unrelated' individuals provides information that constrains the set of possible direct effects, so that including unrelated individuals can increase the effective sample size for estimation of direct effects by up to 50% compared to using samples with genotyped relatives alone. We develop an efficient linear-mixed model approach that uses a sparse genetic relatedness matrix to model relatedness, and is able to estimate both population effects, as in standard GWAS, and direct effects, as in family-based GWAS, using the full sample of individuals, whether they have genotyped relatives or not. Our approach therefore unifies the family-based and standard GWAS approaches. We apply our method to a sample of 408,252 individuals from the UK Biobank, and we estimate the correlation between direct genetic effects and effects estimated by standard GWAS methods for 23 traits, including educational attainment (r=0.738 SE=0.061) and cognitive ability (r=0.699 SE=0.050). Furthermore, direct genetic effects yield lower estimates of SNP heritability (h2) and genetic correlation (rg) than those from standard GWAS estimates for some traits. For educational attainment, SNP h2=0.027 (S.E.=0.013) for direct effects versus 0.141 (S.E.=0.006) for population effects. We show that a unified framework for GWAS that combines nuclear families and unrelated individuals maximizes power for estimating both direct and population effects.

S25. Novel statistical genetics methods for complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 153. Inference of causal networks using bi-directional Mendelian randomization and network deconvolution with GWAS summary data.

Authors:

Z. Lin, H. Xue, W. Pan; Univ. of Minnesota, Minneapolis, MN

Abstract:

Inferring causal relationships among potential risk factors and diseases from observational data is both important and challenging, e.g. due to hidden confounding. Emerging as a powerful tool, Mendelian randomization (MR) has been increasingly applied for causal inference with observational data by using genetic variants as instrumental variables (IVs). However, the current practice of MR has been largely restricted to investigating the total causal effect between two traits, while it would be more useful to infer the direct causal effect between any two of many traits (by accounting for mediating effects through other traits). In this work, we first extend bi-directional MR-cML, a robust MR method based on constrained maximum likelihood, to overlapping-sample MR set-up, then apply it to infer a causal network of total effects among multiple traits. Finally we apply graph deconvolution to infer a causal network of direct effects. Simulation studies showed much better performance of the extended MR-cML with sample overlap. We applied the method to 17 large-scale GWAS summary datasets to infer the causal networks of both total and direct effects among 11 common cardiometabolic risk factors, 4 cardiometabolic diseases (coronary artery disease, stroke, type 2 diabetes, atrial fibrillation), Alzheimer's disease and asthma. The inferred total causal graph identified many well-accepted risk factor-disease pairs while the direct causal graph provided more interesting insights into the mechanisms. We also provide an R Shiny app for users to explore any subset of the 17 traits of interest.

S25. Novel statistical genetics methods for complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 154. A variational Bayesian approach to characterize pleiotropic components across thousands of human diseases and complex traits using GWAS summary statistics

Authors:

Z. Zhang[1], J. Jung[1], S. Gazal[1], N. Mancuso[1,2]; [1]Ctr. for Genetic Epidemiology, Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Dept. of Quantitative and Computational Biology, Univ. of Southern California, Los Angeles, CA

Abstract:

Genome-wide association studies (GWAS) across thousands of traits have revealed the extreme pleiotropy of trait-associated genetic variants, thus providing an opportunity to understand shared disease mechanisms. A recent work proposed to characterize this shared architecture using truncated singular-value decomposition (tSVD) on phenome-wide GWAS summary data. However, this approach does not model effect size uncertainty, making it unclear how traits with different power influence this shared architecture characterization.

To address these limitations, we derive from first principles FactorGo, a factor analysis model on genetic associations, to identify latent pleiotropic factors across thousands of phenotypes using GWAS summary data. FactorGo accounts for uncertainty in effect-size estimates and prunes uninformative factors in a variational Bayes framework. In extensive simulations, we observed that FactorGo outperforms tSVD in capturing latent pleiotropic factors across phenotypes, while maintaining a similar computational cost.

We applied FactorGo and tSVD to estimate 100 latent pleiotropic factors from GWAS summary data at 53,953 non-HLA genetic variants across 2,483 phenotypes measured in European-ancestry Pan-UK BioBank individuals (N=420,531). First, we observed FactorGo identifies phenotype-relevant latent pleiotropic factors that rank focal phenotypes more highly compared with tSVD ($P<2.20e-16$), while capturing intuitive relationships. For example, the leading factor for rheumatoid arthritis (RA) is primarily explained by known RA biomarkers albumin, calcium, RA medication use (methotrexate), in addition to other autoimmune diseases (e.g., inflammatory bowel diseases). Second, we found FactorGo results more effectively span the phenotype spectrum, with the top 10% of traits from pleiotropic factors covering 79% of the total 2483 phenotypes, compared with 73.7% obtained from tSVD ($P=1.23e-5$).

Lastly, we performed an enrichment analysis on factor loadings using 707 cell-type specific annotations across 23 tissues from the IMPACT 707 dataset. We found FactorGo results were more enriched for known phenotype-relevant tissues, compared with those from tSVD. For example, standing height factors were more enriched with heart, bone, muscle, and fibroblast connective tissue annotations ($P=6.25e-5$) and RA factors were more enriched in blood B cells and T cells ($P=2.54e-5$).

Taken together, our results demonstrate that FactorGo prioritizes biologically meaningful latent pleiotropic factors that reflect pervasive shared etiologies across GWAS studies.

S25. Novel statistical genetics methods for complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 155. Fast multiple-trait genome-wide association analysis for correlated longitudinal measurements

Authors:

G. Abdel-Azim, P. Patel, S. Li, S. Guo, M. Black; Janssen Res. & Dev. (Johnson & Johnson), Spring House, PA

Abstract:

Motivation: Large-scale longitudinal biobank data can be used to identify genetic variation contributing to the progression of human diseases and traits. While methods for genome-wide association studies (GWAS) of multiple traits have been proposed, an efficient approach for GWAS of multiple longitudinal traits (LGWAS-MT) is not currently available.
Method: We developed an efficient multiple-trait mixed model approach for GWAS of longitudinal outcomes that includes a fast algorithm for inversion by recursive partitioning of the random effects submatrix. The inverse was used to absorb the large random effects block of the system which significantly reduced the time needed for genome-wide scans. Simulation scenarios for 1,000 to 10,000 individuals, 0.001 to 0.1 of minor allele frequency and 1 to 20% ratio of longitudinal (LT) to cross-sectional (CS) effects were performed to validate algorithms and assess power and type I error. Twenty-eight blood-based quantitative biomarkers measured twice on each participant of the UK Biobank were used to compare method performance and results for standard GWAS on rates of change, (first measure - second measure)/(time between measures), single-trait (LGWAS-ST), and multiple-traits (LGWAS-MT). Genetic association analyses were performed on both the SNP array and whole exome sequencing data available in UK Biobank.
Result: The newly developed inversion algorithm was nearly linear in the number of traits. Consistent with our expectation, power was significantly higher for CS than LT effects, particularly with a diminishing LT/CS ratio. With a minimum minor allele count of 3 within genotype by time categories, observed type I error was roughly equal to theoretical genome-wide significance. Across all biomarkers, we observed 539 (CS) and 249 (LT) significant independent variants for the LGWAS-MT method, and 513 (CS) and 30 (LT) for LGWAS-ST, respectively. Only 37 variants were identified by modeling rates of change, and none of them overlapped with those identified by LGWAS-MT, indicating the low power of rates of change in identifying LT variants. About 98% of CS and 90% of LT independent variants found by LGWAS-ST were also identified by LGWAS-MT across biomarkers analyzed. Number of novel variants found by LGWAS-MT relative to LGWAS-ST was greater for LT than CS components, indicating its advantage in identifying variants contributing to progression.
Conclusion: We developed LGWAS-MT, an efficient multiple-trait mixed model approach for GWAS of longitudinal data of biobank scale, and demonstrated its powerful performance to identify novel associations compared with previously reported methods.

S25. Novel statistical genetics methods for complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 156. Flexibly encoded GWAS identifies novel nonadditive SNPs in individuals of African and European ancestry

Authors:

J. Zhou[1], L. Guare[2], A. Garao Rico[1], T. Gonzalez Zarzar[1], N. Palmiero[1], S. Pendergrass[3], Y. Zhang[4], P. Peissig[5], P. Sleiman[6], H. Hakonarson[6], J. Harley[7,8,9], K. Kiryluk[10], M. Ritchie[2], S. Verma[2], M. Hall[1]; [1]The Pennsylvania State Univ., University Park, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Genentech, South San Francisco, CA, [4]Geisinger Genomic Med. Inst., Danville, PA, [5]Marshfield Clinic Res. Inst., Marshfield, WI, [6]Children's Hosp. of Philadelphia, Philadelphia, PA, [7]US Dept. of VA, Cincinnati, OH, [8]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [9]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [10]Columbia Univ., New York, NY

Abstract:

Genome-wide association studies (GWAS) are the backbone of human genetics research. However, most of the variance for complex diseases remains unexplained. Most GWAS assume an additive inheritance model: heterozygous genotypes are coded with half the risk of homozygous alternate genotypes. Yet, growing evidence indicates that with flexible modeling, many SNPs show nonadditive effects (e.g., dominant and recessive), which will be missed using only the additive model. We developed Elastic Data-Driven Encoding (EDGE) to determine the heterozygous:homozygous alternate ratio of risk. We simulated SNPs with diverse inheritance models (e.g., additive, dominant, recessive), as well as null SNPs, in 1,000 replicates across varying minor allele frequencies (MAFs), sample sizes, case-to-control ratios, and penetrance. Results demonstrated that EDGE outperformed traditional methods (i.e., additive, dominant, recessive, codominant, and dominance deviance) across all simulated models for power ($p<5x10^{-8}$) while maintaining a conserved false positive rate (3.5% to 6.5%). We applied EDGE to biomedical data from the Electronic Medical Records and Genomics (eMERGE) Network for 50,933 age-related macular degeneration (AMD) samples of European ancestry (EUR) and 9,540 AMD samples of African ancestry (AFR) for 5.5 million SNPs. Known regions for AMD were identified (e.g., *CFH* on Chr 1 in EUR). Fifteen of the 60 hits (25%) from both populations were classified by EDGE as nonadditive SNPs. The top result for the AFR subgroup was *rs77408014* (novel for AMD; intron of *CACNA1C*; $p = 1.65 \times 10^{-8}$) with an estimated heterozygous:homozygous alternate risk of 1:4, indicating a sub-additive inheritance model. This research lays the necessary groundwork for integrating nonadditive genetic effects into genetics workflows to improve polygenic risk prediction in diverse populations and springboard future applications to thousands of disease phenotypes, clinical lab measures, biomarkers, and other omic domains to improve disease-prediction capability.

S25. Novel statistical genetics methods for complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 157. fSuSiE: new approach for fine mapping functional phenotypes

Authors:

W. Denault[1], Y. Miao[2], P. Carbonetto[1], G. Wang[1,2], M. Stephens[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Columbia, New York, NY

Abstract:

Fine-mapping genetic variants that influence genome-level processes such as chromatin accessibility and methylation is a key step toward understanding how they influence important organismal-level traits. Such molecular phenotypes can be modeled as a "function" whose value varies continuously along the genome. While many reports show the importance of “functional” quantitative trait loci, it remains difficult to identify which genetic variants regulate these traits.
We introduce a novel Bayesian variable selection method called the “Functional Sum of Single Effects” (fSuSiE) model that extends the SuSiE model of Wang *et al.* (2020) to fine map functional phenotypes such as methylation or chromatin accessibility. fSuSiE is based on a simple nonparametric regression model, and the novelty is in combining this nonparametric regression with a new flexible prior for wavelet regression and in taking an empirical Bayes approach to fit the model efficiently. This enables fSuSiE to be flexible and adaptive to various types of data, such as molecular phenotypes in functional genomic studies. In addition to these features, fSuSiE maintains the key properties of SuSiE, i.e., computational efficiency, intuitive interpretation, and uncertainty quantification of which variables (SNPs) should be selected.
We benchmarked fSuSiE against the best-performing methods, such as ipDMR and Comb-p, for detecting SNPs for genetically differentially methylation regions (gDMR) and genetically comethylated CpGs (gcCpG). We conducted a comprehensive numerical comparison study between fSuSiE, ipDMR, and Comb-p under various scenarios of gDMR and gcCpG. In addition to providing fine-mapped results, we showed fSuSiE is more powerful compared to ipDMR and CombP for detecting gDMRs when the effect of the SNP is modest. Furthermore, fSuSiE largely outperformed other methods for detecting gcCpG.
We applied fSuSiE to identify genetic variants associated with chromatin accessibility (dsQTLs) using data from Degner *et al.* 2012, and methylation (mQTLs) from the Religious order Study and the Memory and Aging Project (ROSMAP) Jager *et al.* 2018. While we restricted our investigation to chromatin accessibility data and methylation data, fSuSiE is also applicable to spatially or temporally structured phenotypes such as longitudinal phenotypes and age-dependent genetic risk factors for disease.

S26. Somatic mosaicism in human health and disease

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 160. Delineating the rates and determinants of human somatic variation among tissues and individuals

Authors:

H. Xu, T. Comi, D. Akey, J. M. Akey; Lewis Sigler Inst. for Integrative Genomics, Princeton Univ., Princeton, NJ

Abstract:

A human zygote undergoes trillions of cell divisions to result in a fully differentiated multicellular organism, providing many opportunities for somatic mutations to occur. Even terminally differentiated cells continue to acquire somatic lesions in their genomes through non-replicative processes, such as spontaneous deamination of methylated cytosines in the context of CpG sites. Thus, the burden of variation in humans is expected to be high, and contribute to many traits and diseases such as cancer, aging, neurodegeneration, and over 30 additional human diseases. However, few systematic and comprehensive studies have been done to characterize levels and patterns of human somatic variability in apparently healthy individuals. To address this important gap in knowledge, as part of eGTEx Project, we performed high-coverage exome sequencing of 496 tissue samples, spanning 46 tissue types, from 23 donors aged between 21 and 69 years. We developed a novel and rigorous statistical framework tailored to the unique structure of our data to identify somatic mutations. As expected, the somatic mutation burden varies widely across tissues and individuals, and is significantly related to age. Epithelial tissues, such as sun-exposed skin and esophagus, are found to have the highest burden of somatic mutations, and mutational signatures related to ultraviolet light are observed from sun-exposed skin. We also find that age is significantly related to the rate of somatic mutations, but only in a a subset of tissues. Finally, we find that many individuals harbor putatively deleterious somatic mutations. Our data provides a high-resolution compendium of somatic mutations across human tissues, which will enable biological insights into mutational mechanisms and help interpret patterns of somatic mutations in the context of disease.

S26. Somatic mosaicism in human health and disease

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 161. Detection, causes and consequences of Y chromosome mosaicism

Authors:

Y. Zhao[1], E. J. Gardner[1], K. A. Kentistou[1], F. Day[1], H. Lango Allen[1], P-R. Loh[2,3], V. Sankaran[4,3,5,6], K. K. Ong[1], J. R. B. Perry[1]; [1]MRC Epidemiology Unit, Inst. of Metabolic Sci., Univ. of Cambridge, Cambridge, United Kingdom, [2]Div. of Genetics, Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Div. of Hematology/Oncology, Boston Children's Hosp., Harvard Med. Sch., Boston, MA, [5]Dept. of Pediatric Oncology, Dana-Farber Cancer Inst., Harvard Med. Sch., Boston, MA, [6]Harvard Stem Cell Inst., Cambridge, MA

Abstract:

Mosaic Loss of chromosome Y (LOY) in leukocytes is the most common form of clonal mosaicism in humans and has a significant germline genetic component. The most recent GWAS to investigate the genetic determinants of LOY in 205,011 males identified 156 independent signals and highlighted a key role for genes involved in cell-cycle regulation and DNA damage response. Population studies typically determine LOY using genotype intensities derived from genotype array data, the accuracy of which varies by the number of Y chromosome probes on the array and are technically noisy. Inaccurate estimation of LOY reduces power to identify genetic and phenotypic associations with LOY. To overcome these constraints, we developed a robust estimator of LOY status derived from several orthogonal approaches using both whole-exome sequence and genotype array data. In genetic data derived from 204,770 UK Biobank men, our new method improved the accuracy of LOY estimation as measured by the strength of association between LOY and age, smoking status, and polygenic risk of LOY derived from previous GWAS. We then used this revised and validated LOY instrument to conduct a new GWAS of LOY status in the UK Biobank. Beyond the previously identified 156 signals, we identified 22 novel LOY-associated loci. Functional enrichment analysis of these genes supports the prior hypothesis that defects in cell-cycle regulation and DNA damage response are major causes of LOY, and also newly implicated sex hormone regulation, notably the Androgen Receptor. We also identified several genes with shared risk for both LOY and cancer, which highlighted the common mechanisms between these two traits. Finally, we leveraged the shared genetic architecture between LOY and other related traits to improve power to identify variants associated with risk of myeloproliferative neoplasm. Based on the current available MPN GWAS summary statistics, we identified 13 novel loci reaching genome-wide significance, including loci near *PARP1*, an established target of cancer therapy. In summary, we show that the accuracy of estimating LOY can be improved by combining multiple approaches using both GWAS array and whole exome sequence data. Our innovative approach improved power to detect novel mechanisms that regulate clonal mosaicism in blood and can be used to enhance the identification of novel genes associated with risk of related cancers.

S26. Somatic mosaicism in human health and disease

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 162. Genetic investigation of mosaic loss of the X chromosome in peripheral leukocytes of 918,085 women identifies germline predisposition and strong signals of haplotype selection

Authors:

A. Liu[1], G. Genovese[2,3], Y. Zhao[4], C. Terao[5], M. Pirinen[1], FinnGen, P-R. Loh[3,6], A. Ganna[1,3,7], J. Perry[4], M. Machiela[8]; [1]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, [2]Stanley Ctr., Broad Inst. of Harvard and MIT, Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Boston, MA, [4]MRC Epidemiology Unit, Inst. of Metabolic Sci., Univ. of Cambridge, Cambridge, United Kingdom, [5]Lab. for Statistical and Translational Genetics, RIKEN Ctr. for Integrative Med. Sci., Tokyo, Japan, [6]Div. of Genetics, Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [7]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [8]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD

Abstract:

Mosaic loss of the X chromosome (LOX) is the most frequently occurring age-related clonal somatic chromosomal alteration detected in peripheral leukocyte DNA of females, with a prevalence of <4% in women aged <50 and reaching >30% after 70. As LOX is a pro-proliferative genomic alteration, understanding the molecular mechanisms driving susceptibility could provide insights into the impact of ageing on hematopoiesis and hematologic cancer risk. GWAS of mosaic loss of the Y chromosome (LOY) in men has identified 156 susceptibility loci (Thompson *Nature* 2019), highlighting genes involved in cell cycle regulation and cancer susceptibility, but no GWAS of LOX has been reported. Here we describe genetic insights on LOX from a meta-analysis of 918,085 women (36,178 with LOX) in 8 biobanks worldwide. We identified 49 independent common susceptibility loci and rare missense variants of *FBXO10* (exome sequencing data, N=226,125) associated with the risk of LOX. Although some LOX loci are shared with LOY (e.g., centromere protein genes *CENPW*, *CENPQ*, *CENPU* and cancer susceptibility genes *TP53*, *MAD1L1*, *ATM*), many loci are specific for LOX, especially HLA alleles (HLA-B*35:01 from MHC class I and HLA-DQB1*04:02 from class II) and 2q37.1 (*SP140L* and *SP110*, immunity's chromatin readers) highlighting relationships of LOX with autoimmunity. The partially distinct genetic underpinnings between LOX and LOY were also supported by moderate genetic correlation (0.30 [0.20-0.40], $P=3x10^{-9}$).
We also performed allelic shift analyses to identify X chromosomes preferentially retained in women with detectable LOX. We identified 45 X chromosome loci that strongly influence which X chromosome is retained when LOX occurs. Multiple variants were identified over a large region spanning the centromere (P<10-650) as well as near chromosome X genes associated with skewed X-inactivation (PLS3, ITM2A), cancer risk/progression (FAM9C, CT45A1, SAGE1), and blood cell counts (P2RY8, WAS, PJA1, PLS3, ITM2A, TMEM255A, SOWAHD). As LOX preferentially involves the inactive X chromosome, such associations support competitive advantage for clonal growth or selection among leukocytes with actively transcribed copies of specific X chromosome loci.
Leveraging genotype data from 918K females, we detect multiple germline variants associated with LOX susceptibility, suggesting relationships with cell cycle regulation, autoimmunity, blood cell traits, and cancer predisposition. Allelic shift analyses further demonstrate a strong *cis* selection of specific X loci, providing evidence of selection for specific loci that could promote skewed X-inactivation and hematologic cancer risk.

S26. Somatic mosaicism in human health and disease

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 163. Meiotic and mitotic aneuploidies drive preimplantation mortality of in vitro fertilized human embryos

Authors:

R. McCoy[1], M. Summers[2,3], A. McCollin[2,3], C. Ottolini[3,2,4], K. Ahuja[2], A. Handyside[3]; [1]Johns Hopkins Univ., Baltimore, MD, [2]London Women's Clinic, London, United Kingdom, [3]Univ. of Kent, Kent, United Kingdom, [4]The Evewell, London, United Kingdom

Abstract:

Preimplantation pregnancy loss is extremely common in humans, both in the contexts of *in vitro* fertilization (IVF) and natural conception. Aneuploidy, the gain and loss of chromosomes either during egg and sperm formation (i.e., meiosis) or postzygotic cell division (i.e., mitosis), has long been recognized as the primary cause of human pregnancy loss. Nevertheless, detailed knowledge of the breadth of chromosome and cell division abnormalities that cause embryonic mortality has been hindered by the fact that preimplantation genetic testing for aneuploidy (PGT-A) is typically restricted to the surviving embryos that are the candidates for transfer after IVF. Here we overcame this bias through comprehensive chromosome analysis (n = 909) and time lapse imaging (n = 843) of a large sample of IVF embryos—both arrested and unarrested—regardless of their morphological grade or other criteria. Across all 909 tested embryos, 206 (22.6%) were deemed euploid while 703 (77.3%) possessed evidence of whole or segmental aneuploidies of one or more chromosomes. Using informative copy number signatures to infer the origins of aneuploidies, we found that both putative meiotic and putative mitotic aneuploidies were enriched among embryos that arrested compared to those that developed to the blastocyst stage, though the effect was much stronger for mitotic (Odds Ratio [OR] = 6.02, $p = 4.22 \times 10^{-33}$) compared to meiotic aneuploidies (OR = 1.62, p = 0.0012). Strikingly, we estimate that 30% of euploid zygotes perished prior to blastocyst formation, largely due to mitotic errors. Moreover, the specific features of aneuploidies, such as the total number of chromosomes affected, predictably influenced the probability of arrest. Abnormal (e.g., multipolar) cell divisions during the initial two embryonic mitoses were prevalent and strongly associated with embryonic arrest (OR = 10.03, $p = 1.04 \times 10^{-42}$), largely via the complex, lethal forms of chromosomal mosaicism that they induced. Meanwhile, among surviving embryos, meiotic and mitotic aneuploidies were strongly associated with impaired blastocyst morphology (X^2 [9, n = 612] = 60.2, $p = 1.23 \times 10^{-9}$). Additionally, the presence of meiotic and/or mitotic aneuploidies was associated with later timing of blastocyst biopsy, indicating that aneuploidies tend to delay the process of blastocyst formation and expansion (Average Marginal Effect [AME] = 0.220, $p = 2.4 \times 10^{-5}$). Together, our study offers a detailed view of the spectrum of aneuploidies in preimplantation embryos and their immense contributions to embryonic mortality, especially during the precarious developmental transition from the cleavage to blastocyst stage.

S26. Somatic mosaicism in human health and disease

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 164. Comprehensive multiomic profiling of somatic mutations in malformations of cortical development

Authors:

C. Chung[1,2], X. Yang[1,2], J. Gleeson[1,2]; [1]Univ. of California San Diego, La Jolla, CA, [2]Rady Children's Inst. for Genomic Med., San Diego, CA

Abstract:

Malformations of cortical development (MCD) are neurological conditions displaying focal disruption of cortical architecture and cellular organization arising during embryogenesis, largely from somatic mosaic mutations. Identifying the genetic causes of MCD has been a challenge, as mutations remain at low allelic fractions in brain tissue resected to treat epilepsy. Here, we report a genetic atlas from 317 brain resections, identifying 69 mutated genes through intensive profiling of somatic mutations, combining whole-exome and targeted-amplicon sequencing with functional validation and single-cell sequencing. Genotype-phenotype correlation analysis elucidated specific MCD gene sets associating distinct pathophysiological and clinical phenotypes. The unique spatiotemporal expression patterns identified by comparing single-nucleus transcriptional sequences of mutated genes in control and patient brains implicate critical roles in excitatory neurogenic pools during brain development, and in promoting neuronal hyperexcitability after birth.

S26. Somatic mosaicism in human health and disease

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 165. Profiling *PIK3CA* variants in disorders of somatic mosaicism

Authors:

B. Mojarad, P. Hernandez, M. Evenson, M. Corliss, M. Schroeder, K. Bowling, J. Heusel, J. Neidich, Y. Cao; Washington Univ. in Saint Louis, Saint Louis, MO

Abstract:

Activating variants in the *PIK3CA* gene upregulate the intracellular PI3K/AKT/MTOR signaling pathway, contributing to uncontrolled cell proliferation and apoptosis resistance. Somatic variants in *PIK3CA* are associated with segmental overgrowth disorders, which are now collectively termed *PIK3CA*-related overgrowth spectrum (PROS). Here we analyze the genetic findings in a cohort of patients with disorders of somatic mosaicism (DoSM) and review the reported, clinically significant variants in *PIK3CA*. Well-characterized recurrent 'hotspots' in *PIK3CA*, such as p.His1047, p.Glu542, and p.Glu545 that are observed across multiple cancer types are also known to drive DoSM; however, in this study we wished to focus on potentially activating *PIK3CA* variants at other positions. A total of 943 individuals with DoSM were referred to our genetics laboratory for next generation sequencing (NGS) between 2013 and 2021. Target enrichment, sequencing and analysis of up to 37 genes known to drive DoSM (including the gene *PIK3CA*) were conducted as previously described (PMID: 31585106). The average depth of coverage was in excess of 2000x, facilitating robust detection of variants at low allelic fractions. In the cohort of 943 individuals with DoSM, a total of 351 reportable *PIK3CA* variants were identified in 350 patients, with two *PIK3CA* variants detected in one individual. The 351 *PIK3CA* variants include 346 clinically significant alterations and five variants of unknown significance. In total, we identified 70 unique variants, of which 12 were small insertions or deletions (indels) and 58 were single nucleotide variants (SNVs). Of the 70 unique variants, six were in the p85 adaptor-binding domain, 25 were in the C2 domain, 12 were in the helical domain, 19 were in the kinase domain, and 8 were in the linker regions between these domains. Notably, we did not find any reportable variants in the RAS binding domain. Importantly, we detected 18 novel pathogenic and likely pathogenic variants (seven indels and eleven SNVs) in patients with PROS, most of which have been previously described in multiple cancer types. Of note, half of these novel variants were identified in the C2 domain. This study analyzed the spectrum of variants in *PIK3CA* using the largest cohort of DoSM-associated clinically significant *PIK3CA* variants from a single center. A group of novel, clinically significant *PIK3CA* variants was identified for the first time in patients with PROS, and together with their correlation with the functional domains, expands the knowledge base of *PIK3CA* variants in DoSM.

S27. Therapeutic development for Mendelian disorders

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 168. Development of a novel oral treatment that rescues gait ataxia and retinal degeneration in a phenotypic mouse model of familial dysautonomia

Authors:

E. Morini[1,2], A. Chekuri[1], E. Logan[1], J. Bolduc[1], E. Kirchner[1], M. Salani[1], A. Krauson[1], J. Narasimhan[3], G. Vijayalakshmi[3], A. Dakka[3], J. Hedrick[3], X. Zhao[3], N. Naryshkin[3], M. Weetall[3], S. Slaugenhaupt[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]PTC Therapeutics, Inc., South Plainfield, NJ

Abstract:

Familial Dysautonomia (FD) is a neurodegenerative disease caused by a splicing mutation in the Elongator complex protein 1 (*ELP1*). This mutation leads to the skipping of exon 20 and a tissue-specific reduction of ELP1 protein, mainly in the nervous system. FD is a complex neurological disorder accompanied by severe gait ataxia and retinal degeneration. Visual acuity typically begins to decline at puberty and often progresses to legal blindness in the third decade of life. Individuals with FD show a significant reduction in the retinal nerve fiber layer (RNFL) due to the death of retinal ganglion cells (RGCs). There is currently no effective treatment to restore ELP1 protein expression in individuals with FD, and the disease is ultimately fatal. After identifying kinetin as a small molecule able to correct the *ELP1* splicing defect, we generated a novel class of compounds that selectively increase the inclusion of exon 20 through the NIH Blueprint Neurotherapeutics Network, and later working with PTC Therapeutics. Here, we optimized the potency, efficacy, and distribution of these compounds to develop an oral treatment that could efficiently pass the blood-brain barrier and correct the *ELP1* splicing defect in the nervous system. We demonstrated that the novel compound, PTC258, efficiently restores correct *ELP1* splicing in mouse tissues, including brain, and most importantly, prevents the progressive neuronal degeneration characteristic of FD. To assess the therapeutic efficacy of PTC258, we treated the phenotypic FD mouse *TgFD9; Elp1*$^{\Delta 20/flox}$ through specially formulated chow. The treatment was well tolerated and improved FD pup survival. We evaluated the effect of PTC258 on mouse gait using CatWalk. Treated FD mice exhibit dose-dependent improvement in motor coordination at 6 months of age, as demonstrated by a progressive increase in both stride length and base of support compared with the vehicle-treated FD mice. This gait improvement correlated with a significant increase in the number of neurons in the dorsal root ganglia (DRG). Similarly, PTC258 rescued retinal degeneration. High-definition spectral-domain optical coherence tomography (OCT) showed a significant dose-dependent improvement in the thickness of the RNFL in the treated FD mice, and retinal flat-mount analysis confirmed the rescue of RGC loss. We show that the phenotypic improvement in mice correlates with the correction of the underlying FD splicing defect, which leads to an increase in full-length *ELP1* transcript and a two-fold increase in functional ELP1 protein in the brain. Our findings highlight the therapeutic potential of this novel class of small molecules as an oral treatment for FD.

S27. Therapeutic development for Mendelian disorders

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 169. High-throughput transcriptome analyses from ASPIRO, a phase 1/2/3 study of gene replacement therapy for XLMTM

Authors:

G. Andreoletti[1], O. Romano[2], H-J. Chou[1], M. J. Sefid-Dashti[1], A. Grilli[2], C. Chen[1], P. Meunier[1], W. Miller[1], F. Mavilio[2], S. Bicciato[2], F. Urbinati[1]; [1]Astellas Gene Therapies, San Francisco, CA, [2]Dept. of Life Sci., Univ. of Modena and Reggio Emilia, Modena, Italy

Abstract:

Introduction: X-linked myotubular myopathy (XLMTM) is a severe congenital disease characterized by profound muscle weakness, respiratory failure, and early death. No approved therapy for XLMTM is available yet. Adeno-associated virus (AAV)-mediated gene replacement therapy (AT132) has shown promise as a novel therapeutic strategy. We aimed to characterize the transcriptomic changes of XLMTM patients from baseline to 24 and 48 weeks after AAV treatment. **Methods**: We leveraged RNA-sequencing data from muscle biopsies of 15 genetically confirmed, ventilator-dependent XLMTM patients, treated with AT132, and enrolled in ASPIRO, a Phase 1/2/3/ clinical trial (NCT03199469). We then applied differential expression analyses, gene co-expression analyses, and machine learning to characterize the muscle transcriptomic changes of these patients before and after AAV treatment. **Results**: RNA-sequencing data indicate that *MTM1* expression levels were significantly (p-value $\leq$0.01) and dose-dependently increased after treatment compared with baseline, as expected. Differential expression analyses between baseline, week 24, and 48 after treatment identified several upregulated genes enriched in lipid metabolism and inflammatory response and down-regulated genes enriched in cell-cell adhesion and muscle development pathways. Further exploration of differentially expressed genes between responders and non-responders based on ventilator dependence identified significantly (p <0.05 and abs log2FC >2) differentially expressed genes involved in inflammatory and immune pathways. Co-expression analysis resulted in similarly regulated genes that were grouped into modules. Finally, random forest classification between baseline, week 24, and 48 identified five genes, including *MTM1*, as potential biomarkers to monitor the response to AAV treatment. These findings further extend our understanding of AT132 in XLMTM1 at a transcriptomics level.

S27. Therapeutic development for Mendelian disorders

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 170. Therapeutic outcomes of selumetinib treatment in pediatric and adult Korean patients with neurofibromatosis type 1 and inoperable plexiform neurofibromas

Authors:

B. Lee[1], H. Kim[1], H. Yoon[1], H. Bae[1], I. Choi[1], H. Do[2], S. Kim[1], J. Kim[1], G-H. Kim[1]; [1]Asan Med. Ctr., Seoul, Korea, Republic of, [2]Asan Inst. for Life Sci., Seoul, Korea, Republic of

Abstract:

Neurofibromatosis type 1 (NF1) is complicated by plexiform neurofibroma (PN), stunted growth, neuropsychiatric difficulties, and reduced quality of life (QOL). Recently, the oral selective MEK inhibitor selumetinib has been reported to decrease the volume of PN, resulting in QOL improvement in pediatric patients in US, EU, and other countries. The current study is an open-label, phase 2 trial, and this interim analysis included the 44 Korean patients, with NF1 (with inoperable, symptomatic or potentially morbid, measurable PN (, treated with selumetinib for at least the first 12 cycles: 10 children at a dose of 20 mg/m^2 q 12hrs, ten children at a dose of 25 mg/m^2 q 12hrs, ten adults at a dose of 50 mg q 12hrs, and 14 adults at a dose of 25mg/m^2 q 12 hrs. Findings of pharmacokinetics, including the maximum concentration (Cmax) and the area under the curve from 0 to 12 hr ($AUC_{0\text{-}12hr}$), volumetric magnetic resonance imaging, growth, neuropsychiatric function, clinical photography of café-au-lait spots (CALS), and QOL questionnaires were evaluated. The most common adverse event (AE) was folliculitis (79.5%), paronychia (34.1%), and pruritus (29.5%). All AEs were CTCAE Grade 1 or 2 and resolved without discontinuation. After the 12 cycles, volume reduction relative to baseline was -35.8% (range, −14.6 to −76.8). Partial responses (tumor volume decreases of 20% or higher from baseline) were confirmed in 42 patients (95.5%). In neuropsychiatric functioning, verbal comprehension (pediatric), perceptual reasoning (adult), processing speed (pediatric and adult), and full-scale IQ (pediatric and adult) scores improved. The height standard deviation score and growth velocity increased in 13 prepubertal patients. The color intensity of CALS decreased in 58.8% of patients after 12 cycles and 83.3% after 26 cycles. The QOL scores improved in 80% of children, 100% of their parents, and 60% of adults. Pain improved shortly following initiation of treatment, and pain duration and frequency gradually decreased duringtreatment. Following a 25 mg/m^2/dose, the mean Cmax and $AUC_{0\text{-}12hr}$ were 1504.8 ng/mL and 2257.4 hr*ng/mL in children and 971.0 ng/mL and 2162.2 hr*ng/mL in adults, respectively, which were higher than in prior Caucasian studies (Eur J Clin Pharmacol 2017;73(6):717-726.). In conclusion, the results of our interim analysis highlights the previously unrecognized, multiple therapeutic effect of selumetinib in both children and adults with NF1 and PN with no serious AE.

S27. Therapeutic development for Mendelian disorders

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 171. Reversible epigenome editing of *PCSK9* as a therapeutic strategy

Authors:

M. Whittaker, I. Jindal, S. V. Cortez-Alvarado, P. Qu, X. Wang, K. Musunuru; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Among the best-established causal risk factors for cardiovascular disease is the blood concentration of low-density lipoprotein cholesterol (LDL-C). Proprotein convertase subtilisin/kexin type 9 (PCSK9), an antagonist to the LDL receptor, has emerged as a promising therapeutic target for the prevention of coronary heart disease. Current PCSK9 inhibitors are administered via subcutaneous injection and their effects are short-lived. An alternative “one-and-done” strategy using genome editing to disrupt PCSK9 at the DNA level has been demonstrated in preclinical animal models, including non-human primates. However, concerns about the permanence and irreversibility of the genomic changes have been raised and might limit the acceptance of the therapies.

Recently, a set of CRISPR-based epigenome editing tools, CRISPRoff and CRISPRon, were reported to regulate gene expression via site-directed methylation and demethylation of gene promoters, respectively. We hypothesized that these epigenome editing tools could durably and reversibly induce methylation changes in the PCSK9 promoter and thereby modulate its expression. We first screened CRISPRoff guide RNAs (gRNAs) targeting the PCSK9 promoter, individually and in dual combinations, for their ability to reduce PCSK9 expression in the human HuH-7 hepatoma cell line. We then performed long-term experiments with the lead candidate gRNAs. We found that these CRISPRoff gRNAs induced profound increases in methylation at CpG dinucleotides in the PCSK9 promoter, with up to 80% decreases in PCSK9 expression. These methylation increases and gene expression decreases have endured through >56 cell divisions so far, with only mild attenuation over time. Using the same gRNAs with CRISPRon in cells previously treated with CRISPRoff, we observed moderate decreases in methylation and increases in PCSK9 expression. Having established these effects in vitro, we are similarly assessing the durability and reversibility of epigenome editing using the lead gRNAs in a PCSK9-humanized mouse model.

Overall, this work provides a proof of concept of precise gene regulation via methylation and demethylation and suggests a potential new therapeutic approach for protection against cardiovascular disease.

S27. Therapeutic development for Mendelian disorders

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 172. Dosage dependent suppression and replacement gene therapy for *KCNH2*-mediated arrhythmias

Authors:

S. Bains, D. Ye, W. Zhou, C. Kim, D. Tester, M. Ackerman; Mayo Clinic, Rochester, MN

Abstract:

Background: *KCNH2*-mediated arrhythmias are caused by either loss-of-function (type 2 long QT syndrome, LQT2) or gain-of-function (type 1 short QT syndrome, SQT1) pathogenic variants in the *KCNH2*-encoded Kv11.1 potassium channel essential for the rapid delayed rectifier current (I_{Kr}) that contributes to the cardiac action potential. No current therapies target the molecular cause of either LQT2 or SQT1. **Objective:** To show a dose-dependent rescue of the pathologic phenotype in patient-specific induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs) of LQT2 and SQT1 using our novel gene therapy. **Methods:** A dual-component "suppression-and-replacement" (SupRep) *KCNH2* gene therapy was created by cloning into a single construct a custom-designed *KCNH2* shRNA that produces ~80% knockdown (suppression) and a "shRNA-immune" (shIMM) *KCNH2* cDNA (replacement). Patient-derived iPSC-CMs and their CRISPR-Cas9 variant-corrected isogenic control (IC) iPSC-CMs were made for a LQT2- (G628S) and a SQT1- (N588K) causative variant. Both pathogenic variant-containing iPSC-CMs were treated with KCNH2-SupRep or a non-targeting control shRNA (sham). Whole cell patch clamp was used to measure the effect of SupRep gene therapy on the action potential duration at 90% repolarization (APD_{90}). **Results:** For the LQT2-causative G628S variant, treatment with KCNH2-SupRep resulted in shortening of the pathologically prolonged APD_{90} to near curative (IC-like) APD_{90} levels (IC, 435±77ms) compared to treatment with sham (G628S: SupRep-treated, 406±147ms vs. sham-treated, 652±132ms, p=0.0009). Conversely, for SQT1-causative N588K variant, treatment with KCNH2-SupRep resulted in therapeutic prolongation of the pathologically shortened APD_{90} (IC: 435±77ms; SupRep-treated: 300±126ms; sham-treated: 161±53ms, p = 0.0052). To account for the incomplete rescue in the SQT1 iPSC-CMs, dose-dependent suppression and replacement was measured using allele-specific qRT-PCR and western blot. Interestingly, robust suppression (>80%) of endogenous *KCNH2* was achieved across all doses. However, replacement with KCNH2-shIMM occurred in a dose-dependent manner thus showing that the expression level of KCNH2-SupRep, and consequently its effect on the APD_{90}, can be adjusted by titrating the therapeutic dose. **Conclusion:** We provide the first proof-of-principle gene therapy for correction of either LQT2 or SQT1. Akin to our sentinel discovery of SupRep gene therapy for LQT1, KCNH2-SupRep gene therapy successfully corrected the pathologic APD_{90}, thereby eliminating the pathognomonic feature of both potentially lethal, *KCNH2*-mediated genetic arrhythmia syndromes.

S27. Therapeutic development for Mendelian disorders

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 10:45 am - 12:15 pm

ProgNbr 173. Base editing as a treatment approach for correction of c.2988+1G to A, the most common CF-causing variant in individuals of African descent

Authors:

E. Kavanagh[1], A. T. Joynt[1], S. Y. Tzeng[1], G. A. Newby[2], N. Sharma[1], D. R. Liu[2], J. J. Green[1], G. R. Cutting[1]; [1]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]The Broad Inst., Cambridge, MA

Abstract:

Alterations of the 5'GT splice motif or 3'AG splice motif result in severe missplicing and are prevalent in many genetic diseases, including cystic fibrosis (CF). While protein-targeted modulator therapies are currently available for treatment of CF, individuals with these canonical splice site variants (CSSVs) are among the ~10% who remain untreated. The most common CF-causing variant in individuals of African descent is a CSSV, c.2988+1G>A. While >75% of individuals with a CSSV have a modulator eligible *in trans* allele, only 50% of the 450 individuals bearing c.2988+1G>A are eligible. Thus, there is a particular unmet need for a treatment for these individuals. CRISPR/Cas9-mediated adenine base editing (ABE) is an efficient and targeted genome editing method correct G>A variants. We electroporated NRCH-ABE8e mRNA and a previously optimized sgRNA to non-differentiated human primary nasal (HNE) or bronchial (HBE) epithelial cells from individuals with CF compound heterozygous for c.2988+1G>A. After differentiation of edited cells genomic editing and recovery of CFTR channel function were assessed. In primary HBEs and primary HNEs, we observed an allelic conversion to WT of 74.7% and 81.3%, respectively, at the +1 site. Interestingly, we also observed high levels of editing at adjacent adenines (+3, +7), which would have a modest effect on mRNA splicing (~20% reduction). However, this did not preclude recovery of CFTR channel function. Compared to WT/WT HBEs and HNEs, unedited cells showed ~5% function, while edited cells achieved >50% function. Since electroporation is not a translationally viable delivery approach, we investigated polymeric nanoparticle mediated delivery to both primary HBE and HNE cells by flow cytometry. GFP mRNA evaluated transfection efficiency and cell viability across three dosages (150, 75, 32.5 ng), four polymer-to-mRNA weight-to-weight ratios (60, 40, 30, 20) and three polymers (R, X, Y). HBEs transfected with polymers R & Y showed ~57% transfection at 75 and 32.5 ng, with polymer X averaging ~25% across the same dosages. Polymer Y showed slightly higher viability of HBEs vs. polymer R. HNEs in comparison achieved a maximum of ~37% GFP transfection with polymers R & Y at 32.5 ng compared to a maximum of ~22% with polymer X at the same dose, with 75-95% viability across all polymers. The ABE design reported here corrects c.2988+1G>A in airway epithelia with high efficiency when robust delivery is achieved. Given that nanoparticle optimization allowed successful delivery to >50% of cells, we anticipate clinically significant recovery of function *in vivo* by combining this ABE design with an optimized polymeric nanoparticle.

S28. Ancestry and admixture in diverse populations

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 176. An ancestry-specific allele frequency browser at 142 million variants and an imputation reference panel derived from exome and whole genome sequencing of the Mexico City Prospective Study

Authors:

A. Ziyatdinov[1], J. Torres[2], J. Alegre-Díaz[3], J. Backman[1], J. Mbatchou[1], M. Turner[4], S. Gaynor[1], T. Joseph[1], Y. Zou[1], D. Liu[5], R. Wade[6], J. Staples[7], R. Panea[1], A. Popov[1], X. Bai[1], S. Balasubramanian[1], L. Habegger[1], R. Lanche[1], A. Lopez[1], E. Maxwell[1], M. Jones[1], H. García-Ortiz[8], &. Ramirez-Reyes[3], R. Santacruz-Benítez[9], A. Nag[10], K. Smith[11], Regeneron Genetics Center, M. Reppell[12], S. Zollner[13], E. Jorgenson[1], W. Salerno[1], S. Petrovski[14], J. Overton[1], J. Reid[1], T. Thornton[1], G. Abecasis[15], J. Berumen[16], L. Orozco-Orozco[17], R. Collins[4], A. Baras[1], M. R. Hill[18], J. R. Emberson[18], J. Marchini[1], P. Kuri-Morales[3], R. Tapia-Conyer[3]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]The Univ. of Oxford, Oxford, United Kingdom, [3]Experimental Res. Unit from the Faculty of Med. (UIME), Natl. Autonomous Univ. of Mexico (UNAM), Mexico City, Mexico, [4]Univ. of Oxford, Oxford, United Kingdom, [5]Regeneron Genetics Ctr., New York, NY, [6]Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [7]Regeneron Genetic Ctr., Tarrytown, NY, [8]Inst. Natl. de Med. Genómica, Tlalpan, Mexico City, Mexico, [9]Experimental Res. Unit from the Faculty of Med. (UIME), Natl. Autonomous Univ. of Mexico (UNAM), Mexico City, Mexico, Mexico City, Mexico, [10]AstraZeneca, London, United Kingdom, [11]AstraZeneca, Melbourn, United Kingdom, [12]AbbVie Inc, North Chicago, IL, [13]Univ. of Michigan, Ann Arbor, MI, [14]AstraZeneca, Fulbourn, United Kingdom, [15]Regeneron Pharmaceuticals, Tarrytown, NY, [16]Univ. Natl. Autónoma de México, México, Mexico, [17]Inst. Natl. de Med. Genómica, Tlalpan, Mexico City, Mexico, Mexico, [18]Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom

Abstract:

Developing genomic studies in diverse populations is a top priority in human genetics. To contribute to these efforts, we generated genotype and exome sequencing data for 140,000 individuals of the Mexico City Prospective Study (MCPS), and whole genome sequencing for 10,000 selected individuals. This resource represents the largest study in Mexico and the most extensive sequenced sample collection of non-European ancestry to date. Whole exome sequencing (WES) and whole genome sequencing (WGS) identified 9.3M (4.0M coding) and 131.9M (1.5M coding) variants respectively, of which 1.4M and 31.5M variants were unique to MCPS compared with other large scale WES and WGS databases.

We developed an imputation reference panel (MCPS10k) from the WGS data, utilizing phase information from sequencing reads and pedigrees. WGS variants were phased onto a genotyping array haplotype scaffold to enable ancestry-specific allele frequency estimation. Through the WGS analysis of trios, we found that haplotypes were phased with an overall switch error rate of 0.0024, but depended upon individual ancestry. We compared the performance of the MCPS10k and TOPMed panels on remaining MCPS and external Mexican individuals, establishing that MCPS10k outperformed TOPMed for variants with MAF > 0.1% and for individuals with greater Native Mexican ancestry. The MCPS10k panel will be available via the Michigan Imputation Server[1].

We devised a novel approach to estimate ancestry-specific allele frequencies that leverages local ancestry (LA) information in genotype data and interpolates ancestry to WES and WGS variants. Median sample sizes for estimation of Native Mexican, European and African ancestry were 91,856, 42,009 and 4,312 respectively for WES variants, and 6,549, 3,058 and 341 for WGS variants. Validating our estimates at MCPS WES variants with those overlapped in gnomAD 3.1, we observed excellent agreement for Non-Finish European ($r^2 = 0.994$) and African ($r^2 = 0.987$) ancestries, despite greater heterogeneity in African populations and the lower median MCPS African sample size. We found the highest levels of variation in African segments consistent across genomic annotations: for example, the mean number of putative loss-of-function variants in Native Mexican, European and African genomes were 347, 361 and 427 respectively. Overall, allele frequencies at 141,802,412 MCPS variants are made publicly available on the RGC website[1], increasing by 10-fold the number of LA-resolved frequencies compared to the gnomAD browser[2].

[1] https://rgc-mcps.regeneron.com/home[2] https://gnomad.broadinstitute.org/[3] https://imputationserver.sph.umich.edu/

S28. Ancestry and admixture in diverse populations

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 177. IBD-Mix: An accurate IBD segment-based local ancestry inference method

Authors:

D. Zhi[1], Y. Wei[2], S. Zhang[2]; [1]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Univ. of Central Florida, Orlando, FL

Abstract:

With the increasing popularity of whole genome sequencing technology and diverse biobank cohort, the high resolution of genetic structure of admixed populations is to be discovered. Identifying the ancestry of each segment of an individual haplotype, or local ancestry inference (LAI), is the key process to reveal such structure. Current competitive LAI methods typically use either site-level statistical models or window-level machine learning models. All these methods do not leverage the intuition that local ancestry breakpoints are also identical-by-descent (IBD) segment breakpoints as they are all caused by recombination events. With the availability of large panels for reference populations and the fact that the IBD segments can be called efficiently, here, we present a new method, IBD-Mix, to perform local ancestry inference by leveraging the power of Identity-By-Descent (IBD) segments. IBD-Mix first calls IBD segments between the query haplotype and template haplotypes in the reference panel by RaPID-Query, and then builds a graph with different weights of switching templates within or between ancestries. The LAI can be formulated as an optimization problem by finding the shortest path in the graph that can be solved efficiently by dynamic programming. For benchmarking, admixed individuals are simulated by SLiM using 1000 Genomes project dataset as founders' populations. Initial results suggest that for a 15-generation two-way admixed population (African and European), the accuracy of IBD-Mix reaches 99.72%, which outperforms the state-of-the-art LAI method G-Nomix having 98.69%. For a 15-generation three-way admixed population (African, European, and East Asian), the accuracy reaches 99.00% as G-Nomix has 95.04%. Since IBD-Mix outperforms G-Nomix which was reported outperforming RFmix and other competitive methods, we expected that IBD-Mix will be a competitive method for advancing genetics studies of admixed populations.

S28. Ancestry and admixture in diverse populations

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 178. Using topological ancestry groups to describe population structure in large multi-ethnic biobanks

Authors:

A. Diaz-Papkovich, S. Gravel, S. Zabad, V. Nathan, L. Anderson-Trocme, C. Ben-Eghan, J. Patel; McGill Univ., Montreal, QC, Canada

Abstract:

Population structure is ubiquitous in genetics as a confounder in genome wide association studies. Population genetic analyses typically focus on identifying major axes of variation (PCA) or ancestral components within individuals (Admixture). However, individuals from populations that share demographic histories can vary appreciably in PCA or admixture components. Here, we instead use a topological approach to identify groups with similar genetic ancestries and demonstrate its effectiveness at identifying different scales of structure in large, multi-ethnic biobanks with unbalanced population sizes and many admixed individuals.

Using genotype data, we apply UMAP with HDBSCAN(e), an unsupervised hierarchical density-based clustering algorithm for varying population sizes. Clustering of the 1000 Genomes data strongly matched population labels, with a median Rand index of 0.94. In the UK biobank (UKB), this approach creates clusters correlated to place of birth and self-identified ethnic background, as well as clusters not easily captured by questionnaire data, providing a novel method for identifying populations in a label-agnostic manner. These topology-based clusters show distinct distributions of phenotypic measures, environmental data, and socioeconomic data.

This approach identifies unique phenotype distributions in populations that persist even after correcting with top principal components, such as in populations with recent admixture. It is computationally cheap, simple to integrate into linear regressions and linear mixed models, and is well-suited to large multi-ethnic biobanks. We discuss implications for variant discovery, for the construction and transferability of polygenic risk scores across ethnic groups, and for the identification of environmental interactions.

S28. Ancestry and admixture in diverse populations

Location: Conv Ctr/West Hall A/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 179. On the number of genealogical ancestors tracing to the source groups of an admixed population

Authors:

J. Mooney[1], L. Agranat-Tamir[2], J. Pritchard[2], N. Rosenberg[2]; [1]Univ. of Southern California, Los Angeles, CA, [2]Stanford Univ., Stanford, CA

Abstract:

In a genetically admixed population, admixed individuals possess ancestry from the various source groups. Studies of genetic admixture in admixed human populations frequently estimate ancestry components corresponding to fractions of individual genomes that trace to specific ancestral populations. However, the same numerical ancestry fraction can represent a wide array of admixture scenarios. Using a mechanistic model of admixture, we characterize admixture genealogically: how many distinct ancestors from the source populations does the admixture represent? We consider the case of African Americans in the United States, for whom estimates of continent-level ancestry produce a 75-85% value for African ancestry on average and 15-25% for European ancestry. Genetic studies together with numerical values drawn based on key epochs of African-American demographic history suggest ranges for model parameters. Using the model, we infer that if the genealogical lineages of a random African-American individual are traced back until they reach members of source populations, the expected number of genealogical lines terminating with African individuals is approximately 300, and the expected number terminating in Europeans is approximately 50. This genealogical perspective on admixture can contribute to further understanding the processes that underlie current admixed populations. Moreover, for African Americans, the results provide insight both on how many of the ancestors of a typical African American might have been forcibly displaced to the Americas in the Transatlantic Slave Trade and on how many separate European admixture events might exist in a typical African-American genealogy.

S29. Emerging topics in biobank-scale association analysis

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 182. Genetic analysis of breast density phenotypes derived using deep learning applied to the UK Biobank MRI dataset

Authors:

B. J. Geraghty[1], M. D. Kessler[1], K. Landheer[2], M. Germino[2], Regeneron Genetics Center, J. R. Walls[2], E. Jorgenson[1], J. L. Marchini[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Regeneron Pharmaceuticals, Tarrytown, NY

Abstract:

Mammographic breast density (BD) is well-established as one of the strongest independent risk factors for breast cancer. Dense breast tissue is comprised of fibroglandular tissue, whereas nondense tissue is mainly fat. Women with mammographic density in 75% of the breast have a 4- to 5-fold increased risk of breast cancer as compared to those with little or no dense tissue. BD is a highly heritable, quantifiable trait, and several recent genome wide association studies (GWAS) have identified over 50 loci associated with mammographic BD phenotypes. Many of these variants have been discovered as the susceptible loci of breast cancer; therefore, uncovering the genetic basis of BD may help identify genes that determine breast cancer risk. In contrast to mammography, MRI can provide volumetric measures of BD that are potentially more sensitive to individual variability. For the first time, we have derived BD phenotypes from MRI scans of 15,614 women from the UK Biobank (UKB) cohort, and show that these lead to replication of existing genetic associations for BD, and novel discoveries, paving the way for future genetic studies of MRI-derived BD. We obtained whole-body water and fat MRI scans from UKB and trained a convolutional neural network to perform breast volume segmentation using manual annotations from 122 participants. Water- and fat-fraction maps were averaged within the breast volume segmentations to derive the dense volume (DV) and nondense volume (NDV) phenotypes, respectively. Percent dense volume (PDV) phenotypes were taken as the proportion of DV to DV+NDV. GWAS of DV, NDV and PDV was performed to test for genetic signals known to associate with mammographic BD. We identified 15 genome-wide significant signals, including 11 that recapitulated known mammographic BD loci, and novel associations with DV at loci overlapping the *EGFR* and *ROR1* genes. We also replicated 39/57 signals previously identified to associate with mammographic BD. As the UKB imaging cohort grows to 100,000 participants and beyond, this work could yield the single largest sample of BD phenotypes and facilitate rare variant discovery. These phenotypes will be returned to UKB, providing an important resource for investigating the genetic basis of breast cancer.

S29. Emerging topics in biobank-scale association analysis

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 183. Age-dependent topic modelling of comorbidities in UK Biobank identifies disease subtypes with differential genetic risk

Authors:

X. Jiang[1], Y. Zhang[2], M. Zhang[3], C. Holmes[1], A. L. Price[4], G. McVean[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Big Data Inst., Oxford, United Kingdom, [3]Harvard Univ., Cambridge, MA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Longitudinal data from electronic health records (EHR) has immense potential to improve clinical diagnoses and personalized medicine (Abul-Husn et al. 2019 Cell). A key challenge is to extract biologically useful information from age-dependent patient disease histories spanning a large number of distinct diseases.
We introduce an age-dependent topic modelling (ATM) method that provides a low-rank representation of longitudinal records of hundreds of distinct diseases in large EHR data sets. The model learns, and then assigns to each individual, topic weights for several disease topics, each of which reflects a set of diseases that tend to co-occur as a function of age (quantified by age-dependent topic loadings for each disease). We use variational Bayesian methods to estimate the parameters of the model. Simulations show that ATM outperforms other approaches in distinguishing distinct age-dependent comorbidity profiles.
We applied ATM to 282,957 UK Biobank samples, analyzing 1,726,144 disease diagnoses spanning 349 diseases with ≥1,000 incidences. We identified 10 disease topics optimizing model fit. Most disease topics grouped diseases from the same system (e.g. cardiovascular or endocrine) and most diseases were assigned to a single topic. However, we identified 52 diseases with heterogeneous comorbidity profiles (≥500 incidences assigned to each of ≥2 topics), including breast cancer, type 2 diabetes (T2D), hypertension, and hypercholesterolemia; for most of these diseases, the topic loadings were highly age-dependent, implying differences in disease etiology for early-onset vs. late-onset disease. We note that these analyses were not informed by genetic data.
We defined subtypes of the 52 heterogeneous diseases based on their respective topics, and sought to validate the disease subtypes by comparing genetic risk across subtypes using polygenic risk scores (PRS) (computed in held-out samples using LD-pruning + P-value thresholding). We identified several diseases whose PRS differed significantly across subtypes, including the cardiovascular subtype of T2D (0.20 s.d. higher PRS; p=6e-10) and the sensory subtype of asthma (0.07 s.d. higher PRS; p=8e-7). We further identified specific SNPs underlying these differences. For example, the T2D-associated SNP rs1412829 in the *CDKN2B* locus has a higher odds ratio in the top quartile of cardiovascular topic weight (1.15±0.02) than in the bottom quartile (1.03±0.02) (p=4e-05 for difference). These results validate the biological significance of the disease subtypes inferred using ATM.

S29. Emerging topics in biobank-scale association analysis

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 184. Refined patient stratification leveraging biomarker and other quantitative traits from UK Biobank to enhance PheWAS analyses

Authors:

M. Garg[1], M. Karpinski[1], I. Melas[1], A. O'Neill[1], Q. Wang[2], A. Harper[3], S. Petrovski[1], **D. Vitsios**[1]; [1]AstraZeneca UK Ltd, Melbourn, United Kingdom, [2]AstraZeneca Pharmaceuticals LP, Waltham, United Kingdom, [3]AstraZeneca UK Ltd, Cambridge, United Kingdom

Abstract:

Refined patient stratification and sufficient sample sizes are critical factors in identifying confident genetic signals from phenome-wide association studies (PheWAS). Characterising patient cohorts with high confidence can suffer from incompleteness of electronic health records, inaccuracies in self-reported diseases, undiagnosed conditions due to an early stage of the disease at the time of assessment and other factors. UK Biobank (UKB) offers a wealth of phenotypic information that can be used to tackle these challenges and discover previously undiagnosed subjects by predicting disease likelihood from quantitative traits and biomarker profiles. Leveraging this information, we can build patient cohorts of higher confidence and with greater sample sizes, thus improving the power of PheWAS analyses and uncovering novel signals. Herein, we present MILTON (MachIne Learning with phenoType associatONs): a framework for phenome-wide inference of disease signatures using machine learning. The learnt models enable the construction of extended case cohorts that can be used for further analyses, such as rare-variant collapsing analyses, as well as estimate the importance of key factors contributing to each disease. We applied MILTON to 5,143 ICD10-based cohorts with a sufficient number of known cases (> 300), achieving an AUC > 0.80 for 340 of those phenotypes. The trained models infer the signature of each phenotype or disease based on a set of 58 quantitative traits (mostly biomarkers) available in UKB and then extract novel cases from the rest of the UKB cohort, predicted to have similar biomarker profile with the original disease cohort. By performing binary PheWAS analysis on MILTON extended cohorts, we were able to pick up 301 out of 534 significant associations ($p < 5x10^{-8}$), previously missed by binary PheWAS collapsing analysis but picked up by quantitative PheWAS ($p < 5x10^{-8}$) in our published PheWAS study on 300K UKB WES samples (Wang et al., 2021). We also found 1,193 completely novel hits of higher confidence (MILTON AUC > 0.90) as well as thousands more that require further investigation. We accompany MILTON's phenome-wide results with a web-based portal that enables: 1) exploration of novel associations predicted by MILTON, 2) inspection of top contributing features learnt for each phenotype and 3) comorbidity analysis across ~13,500 ICD10 codes available in UKB for any custom case cohort to help with the interpretation of the results. We find that MILTON's results provide novel insights into the profile of several diseases as well as elucidating underpowered gene-disease associations across multiple phenotypes.

S29. Emerging topics in biobank-scale association analysis

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 185. Participation bias correction reveals substantial impact on genetic association and downstream analyses in the UK Biobank

Authors:

T. Schoeler[1], D. Speed[2], E. Porcu[1], N. Pirastu[3], J-B. Pingault[4], Z. Kutalik[1]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Aarhus Univ., Aarhus, Denmark, [3]Human Technopole, Milan, Italy, [4]Univ. Coll. London, London, United Kingdom

Abstract:

While large-scale volunteer-based biobanks have become the cornerstone of genetic epidemiology, such samples are rarely representative of their target population. For example, of the 9 million people invited to take part in the UK Biobank (UKBB), only 5.5% (~500,000) were recruited into the study - a sample of volunteers with healthier lifestyles, higher levels of education, and favourable socio-economic conditions compared to the general population. The extent to which selective participation in the UKBB distorts findings obtained in genetic studies is currently unknown. In this study, we aim to evaluate the impact of participation bias in the UKBB, and to pin down areas of research that are particularly susceptible to bias when using non-representative samples for genome-wide discovery. By comparing 11 harmonised characteristics of the UKBB to that of a representative sample, we derived a model for participation probability. We then conducted inverse probability weighted genome-wide association analyses (wGWA) on more than 20 UKBB traits, and compared results to those obtained from standard GWA analyses. Comparing the output obtained from wGWA ($N_{effective}$=72,000-93,000) to traditional GWA analyses (N=237,000-303,000), we assessed the impact of participation bias on three sets of genomic findings, namely 1) genome-wide discovery, 2) SNP heritability and genetic correlation estimates and 3) exposure-outcome associations obtained from Mendelian Randomization (MR). We find that about 3% of all identified SNPs reached genome-wide significance only in wGWA analyses, highlighting SNPs missed as a result of participation bias. Participation bias can lead to both an over-estimation (e.g., for education) and under-estimation of SNP effects (e.g., smoking). In a similar vein, participation bias distorted heritability estimates (average change: 0.029, range 0.001-0.258), genetic correlations (average change: 0.054, range 0.001-0.264) and MR estimates (up to 0.31 change in $b_{standardized}$), most notably for socio-behavioural traits including education, smoking and BMI. Overall, the bias mostly affected magnitude of effects, rather than direction. In contrast, genome-wide findings for more physical health traits (e.g., LDL, SBP) exhibited less bias as a result of selective participation. In summary, our results highlight that participation bias can distort genomic findings obtained in non-representative samples. Moving forward, more efforts ensuring either sample representativeness or correcting for participation bias are paramount, especially when studying the genetic underpinnings of behaviour, lifestyles and educational outcomes.

S30. Genotypes and phenotypes of Mendelian disorders

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 188. The inner junction protein CFAP20 functions in motile and non-motile cilia and is critical for vision

Authors:

P. Au[1], P. W. Chrystal[2], N. J. Lambacher[3], L. P. Doucette[4], J. Bellingham[5], E. R. Schiff[5,6], N. C. L. Noel[7], S. Tsiropoulou[8], G. A. Casey[7], Y. Zhai[4], N. J. Nadolski[7], M. H. Majumder[2], J. Tagoe[9], F. D'Esposito[10,11], M. F. Cordeiro[11], S. Downes[12], J. Clayton-Smith[13], J. Ellingford[14], Genomics England Research Consortium, O. A. Mahroo[5,6], J. Hocking[7], M. E. Cheetham[5], A. Webster[5,6], G. Jansen[15], O. E. Blacque[16], W. T. Allison[2,7], I. MacDonald[7,4], G. Arno[5,6,17], M. Leroux[3]; [1]ACHRI, Univ. of Calgary, Cumming Sch. of Med., Calgary, AB, Canada, [2]Dept. of Biological Sci., Univ. of Alberta, Edmonton, AB, Canada, [3]Dept. of Molecular Biology and Biochemistry, and Ctr. for Cell Biology, Dev., and Disease, Simon Fraser Univ., Burnaby, BC, Canada, [4]Dept. of Ophthalmology & Visual Sci., Univ. of Alberta, Edmonton, AB, Canada, [5]UCL Inst. of Ophthalmology, London, United Kingdom, [6]Moorfields Eye Hosp., London, United Kingdom, [7]Dept. of Med. Genetics, Univ. of Alberta, Edmonton, AB, Canada, [8]Sch. of Biomolecular and BioMed. Sci., Conway Inst., Univ. Coll. Dublin, Belfield, Dublin, Ireland, [9]Lethbridge Outreach Genetics Service, Alberta Hlth.Services, Lethbridge, AB, Canada, [10]Eye Clinic, Dept. of NeuroSci.s, Reproductive Sci. and Dentistry, Univ. of Naples Federico II, Naples, Italy, [11]Imperial Coll. Ophthalmic Res. Unit, Western Eye Hosp., Imperial Coll. Hlth.care NHS Trust, London, United Kingdom, [12]Oxford Eye Hosp., Oxford Univ. Hosp. NHS Fndn. Trust, Oxford, United Kingdom, [13]Manchester Ctr. for Genomic Med., St Mary's Hosp, Manchester, United Kingdom, [14]Univ. of Manchester, Manchester, United Kingdom, [15]Dept. of Cell Biology, Erasmus Univ. Med. Ctr., Rotterdam, Netherlands, [16]Sch. of Biomolecular and BioMed. Sci., Conway Inst., Univ. Coll. Dublin, Belfield, Ireland, [17]North Thames Genomic Lab. Hub, Great Ormond Street Hosp. for Children NHS Fndn. Trust, London, United Kingdom

Abstract:

Genetic disorders affecting cellular cilia are known as ciliopathies, and are typically classified based on if the genetic alteration impacts motile or non-motile sensory cilia. Motile cilia dysfunction are associated with primary ciliary dyskinesia, characterized by chronic respiratory airway disease. In contrast, defects of non-motile cilia cause variable phenotypes such as retinopathy, renal disease, and neurological disorders. CFAP20 is a highly conserved protein across multiple organisms with an established role in motile cilia in unicellular eukaryotes. We report 8 individuals from 4 unrelated families who have inherited retinal dystrophy (IRD) due to biallelic variants in *CFAP20*. 3 families were sequenced as part of an undiagnosed retinal dystrophy cohort via 100K Genomes project, and the fourth family was identified through clinical exome sequencing. 3 families presented with isolated IRD but there was a potentially more severe phenotype in family 4 where there were variable neurological phenotypes as well (epilepsy, spasticity, and neurodevelopmental disorder). No individuals had respiratory airway disease. All variants were rare in gnomAD v2.1.1 with MAF of 1.06E-05 or less. The affected individuals were either homozygous for a missense variant, or compound heterozygous for a missense variant with a splice or frameshift variant, and notably no individuals were biallelic for loss of function variants. All missense variants were predicted damaging, and CFAP20 demonstrates constraint for missense variation (Z=2.07). The patient specific CFAP20 protein variants expressed in HEK cells had decreased stability compared to WT controls. In zebrafish, we found that CFAP20 is required for motile cilia function. However, zebrafish *cfap20*-/- mutants also have a progressive retinal dystrophy reflecting the human phenotype. Of note, unlike the other missense variants, the Y86C variant from family 4 was not able to rescue *cfap20*$^{-/-}$ zebrafish, suggesting this variant may be more damaging. In *C. elegans*, an organism without motile cilia, CFAP20 maintains the structural integrity of the inner junction of non-motile cilia and influences behaviours and development. This report links IRD, a classic non-motile ciliopathy phenotype, to a gene with a known role in motile cilia, and thus challenges the motile vs non-motile ciliopathy dichotomy. More families will be needed to clarify the phenotypic spectrum of CFAP20-related disease, as syndromic disease with systemic / neurological manifestations may be possible as suggested by family 4, and different variants may be associated with variable functional impairment of the protein.

S30. Genotypes and phenotypes of Mendelian disorders

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 189. Defining the clinical and molecular spectrum of a *KDM6B*-related neurodevelopmental disorder through large cohort analysis, 3D-protein structure analysis, and *Drosophila* functional assays

Authors:

T. Jakub[1], D. Rots[2], A. Jackson[3], C. Keung[4], S. Banka[3], V. Ibrahim[1], D. Top[1], F. Di Cara[1], R. J. Louie[5], E. S. Stolerman[5], H. Brunner[2], L. Vissers[2], J. M. Kramer[1], T. Kleefstra[2]; [1]Dalhousie Univ., Halifax, NS, Canada, [2]Radboudumc, Nijmegen, Netherlands, [3]Univ. of Manchester, Sale, United Kingdom, [4]Western Univ., London, ON, Canada, [5]Greenwood Genetic Ctr., Greenwood, SC

Abstract:

Development of the brain requires precise regulation of gene expression. Epigenetics plays a critical role in gene expression regulation and disruption of epigenetic regulators frequently results in neurodevelopmental disorders (NDDs). KDM6B is a histone demethylase that specifically removes methyl groups from di- and trimethylated histone H3 lysine 27 (H3K27me3), a well-established epigenetic modification for gene silencing. Previously, 12 *de novo KDM6B* variants were found to be associated with a dominant syndromic NDD. However, the phenotypic spectrum of the disorder remains undefined and the underlying molecular mechanisms are still unknown. Here, we vastly expand the molecular and clinical spectrum of the KDM6B-related NDD by analyzing 85 new cases with mostly *de novo KDM6B* variants. Combining our novel cohort with the previously published cases, we found a total of 82 unique variants in *KDM6B*: 61 loss of function (LoF) and 21 protein altering variants (PAVs). Using *in silico* 3D protein structure analysis and a novel dual *Drosophila* gain-of-function assay, we demonstrated a disruptive effect of 11 PAVs located in or near the enzymatic JmJC domain or Zn-containing domain of KDM6B. The significance of 10 PAVs remains unknown. Individuals with confirmed likely pathogenic *KDM6B* variants present with neurodevelopmental issues, along with other phenotypes with varying prevalence and severity. Analyzing a large cohort allowed for identification of important but rare features of the KDM6B-related NDD including seizures or psychosis present among older individuals in the study. While the overall phenotype is too variable to be recognized by clinical assessment alone, cognitive deficits are seen consistently in all individuals. We therefore assessed the role of KDM6B in cognition and behaviour using a neuronal *Drosophila* RNAi knockdown model, which shows deficits in memory, social behaviour, sleep, and activity. Taken together, we define the molecular and clinical spectrum of the KDM6B-related NDD, introduce a novel functional testing paradigm for the assessment of *KDM6B* PAVs, and demonstrate a conserved role for KDM6B in cognition and behaviour.

S30. Genotypes and phenotypes of Mendelian disorders

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 190. *De novo* variants in *MAST4* cause a neurodevelopmental disorder with variable brain malformations and epilepsy

Authors:

S. Houghtaling[1], A. Vilceanu[2], K. Gripp[3], D. Doherty[4], M. Ruzhnikov[5], K. Sisco[6], D. Wenger[6], Undiagnosed Disease Network, S. Sacharow[7], M. Osmond[8], Care4Rare, K. Lanko[9], A. Bouman[9], D. Beier[1], S. Barakat[9], D. A. Keays[2], K. Aldinger[1]; [1]Seattle Children's Res. Inst., Seattle, WA, [2]Ludwig-Maximilians-Univ. Munich, Munich, Germany, [3]Nemours Children's Hosp., Wilmington, DE, [4]Univ of Washington, Seattle, WA, [5]Lucile Packard Children's Hosp., Stanford, CA, [6]Washington Univ. Sch. of Med., St. Louis, MO, [7]Boston Children Hosp., Chestnut Hill, MA, [8]Children s Hosp. of Eastern Ontario, Ottawa, ON, Canada, [9]Erasmus Univ. Med. Ctr., Rotterdam, Netherlands

Abstract:

The MAST family of microtubule-associated serine-threonine kinases has been implicated in neurodevelopmental disorders, including developmental brain abnormalities (MAST1) and epilepsy (MAST3). Through an international collaboration, we identified 7 individuals with *de novo* heterozygous variants in *MAST4*, including one recurrent missense variant associated with a consistent brain malformation in two unrelated patients. All *MAST4* patients presented with developmental delay and vision abnormalities. Neuroimaging findings included periventricular leukomalacia, cerebellar atrophy, polymicrogyria, and mega corpus callosum. A developmental and epileptic encephalopathy was diagnosed in 4 individuals. To gain insight into the role of *MAST4* in neurodevelopment, we performed *in situ* hybridization in the human brain during early developmental stages and found that *MAST4* is prominently expressed in the developing thalamus. To investigate the effect of MAST4 loss-of-function *in vivo*, we used CRISPR/Cas9 to efficiently reduce *mast4* expression in developing zebrafish embryos and quantified head size at 48 hours post fertilization. Head volume was reduced relative to the yolk sac volume in *mast4* injected zebrafish, suggesting abnormal brain development due to reduced *mast4* levels. Finally, to determine whether MAST4 mutations influence microtubule binding, we are performing *in vitro* transcription and translation assays. In summary, we identified 7 individuals with *de novo MAST4* variants with overlapping neurological phenotypes. Our initial analysis in a zebrafish model suggests that mast4 function is required for proper brain development.

S30. Genotypes and phenotypes of Mendelian disorders

Location: Conv Ctr/Petree D/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 191. Standardized phenotypic similarity analysis in 10,705 exome sequencing trios with 619,109 clinical annotations reveals hidden patterns in neurodevelopmental disorders

Authors:

S. Ganesan[1,2,3,4], S. Parthasarathy[1,2,3], S. M. Ruggiero[1,2,3], P. D. Galer[1,2,3,5], D. Lewis-Smith[6,7,2,1], S. R. Cohen[1,3], M. Pendziwiat[8], L. Smith[9], H. C. Mefford[10], A. Poduri[11,12,13,14], I. Helbig[1,2,3,15]; [1]Div. of Neurology, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of BioMed. and Hlth.Informatics (DBHi), Children's Hosp. of Philadelphia, Philadelphia, PA, [3]The Epilepsy NeuroGenetics Initiative (ENGIN), Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Sch. of BioMed. Engineering, Sci. and Hlth.Systems, Drexel Univ., Philadelphia, PA, [5]Univ. of Pennsylvania, Ctr. for Neuroengineering and Therapeutics, Philadelphia, PA, [6]Translational and Clinical Res. Inst., Newcastle Univ., Newcastle, United Kingdom, [7]Royal Victoria Infirmary, Newcastle-upon-Tyne, Newcastle, United Kingdom, [8]Inst. of Clinical Molecular Biology, Christian-Albrechts-Univ. of Kiel, Kiel, Germany, [9]Epilepsy Genetics Program, Dept. of Neurology, Boston Children's Hosp., Boston, MA, [10]Ctr. for Pediatric Neurological Disease Res., St. Jude Children's Res. Hosp., Memphis, TN, [11]Dept. of Neurology, F.M. Kirby Neurobiology Ctr., Boston Children's Hosp. - Harvard Med. Sch., Boston, MA, [12]Epilepsy Genetics Program, Dept. of Neurology, Boston Children's Hosp. - Harvard Med. Sch., Boston, MA, [13]Dept. of Neurology, Harvard Med. Sch., Boston, MA, [14]Div. of Epilepsy and Clinical Neurophysiology, Dept. of Neurology, Boston Children's Hosp., Boston, MA, [15]Dept. of Neurology, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA

Abstract:

Next generation sequencing (NGS) has helped us to identify disease-causing *de novo* variants that are associated with various neurodevelopmental disorders. Diagnosing these variants have been arduous since genetic etiologies often present with a wide range of phenotypic heterogeneity. Here, we harmonized and analyzed 10,705 trios with exome sequencing data across four cohorts and translated 103,523 specified clinical features to Human Phenotype Ontology (HPO) terminology. The clinical annotations spanned across 4,357 terms including global developmental delay (27%), and seizures (12%) as the most common terms. After inferring higher-level terms, we determined a total of 619,109 clinical annotations including 5,366 unique terms. Analyzing using a standardized variant calling pipeline and filtering for *de novo* variants, we diagnosed 1,183 variants and 298 genetic etiologies present in two or more individuals. *TRIM13* (n=33), *KMT2A* (n=22), *KCNQ2* (n=18), *MECP2* (n=17), and *STXBP1* (n=17) were the most common genetic etiologies in the cohort. We also identified 102 recurrent *de novo* variants in two or more individuals, with *TRIM13* p.L407F (n=33), *PACS1* p.R203W (n=15), and *PPP2R5D* p.E92K (n=12) representing the most frequent recurrent variants. Leveraging the structure of HPO, we measured clinical relatedness using phenotypic similarity analysis and identified 72 genes with significant phenotypic similarity. For eight genetic etiologies that did not reach significance, recurrent variants in these genes showed significant similarity including *ATN1* p.Q496_502del, and *SLC6A1* p.F287S. Clustering analysis revealed seven genes including *STXBP1* and *BPTF* with subgroups characterized by distinct phenotypic features. We also compared phenotypic similarity with the expected frequency of *de novo* variants, highlighting patterns in phenotypically similar genetic etiologies. Using a novel phenotypic neighborhood approach, we examined individuals with high phenotypic similarity to the carriers of *de novo* variant in the same gene. Most genetic etiologies only accounted for a fraction of clinical similarity between individuals, highlighting the complex phenotypic landscape of neurodevelopmental disorders. In summary, we provide a scalable framework for phenotype analysis in large datasets to improve delineation of gene-disease relationships. Using computational phenotypes methods for the ever-increasing cohorts of sequencing data along with rich phenotypic data, this provides the possibility to detect novel genetic etiologies especially when the individuals have a rare causative gene with prominent phenotypic features.

S31. How do we express ourselves? Examining promoter and enhancer biology

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 194. Allelic specific transcription factor binding and chromatin accessibility shape promoter kinetics in human cell lines

Authors:

B. Jin, W. Bush; Case Western Reserve Univ., Cleveland, OH

Abstract:

Bulk RNA-seq expression is an average measurement between two chromosomes and across cell populations, which diminishes the allelic and population gene expression heterogeneity. Mechanistically, gene expression is controlled by promoter kinetics, which typically produces a bursting pattern. These bursting patterns create significant transcriptional heterogeneity in human cells and are associated with cell differentiation, aging, and more. Despite their importance, promoter kinetics have not been comprehensively studied in human cells. We used single-cell RNA-seq from two lymphoblastoid cell lines (LCLs) to achieve a transcriptomic measurement of promoter kinetics (promoter switching rates, mRNA transcription rate, burst size, and burst frequency). We combined scRNA-seq data with phased genotypes and fit them with a two-state model to derive promoter kinetics for a single chromosome. We derived 161 allelic-level transcription factors (TFs) occupancy profiles from ChIP-seq data and associated the allelic-specific binding (ASB) of TFs with allelic-level promoter kinetics. We further validated the results with allelic open chromatin derived from DNA methylation and ATAC-seq profiles. In GM12878 and GM18052, over 90% of biallelic expression observed in the cell population comes from random monoallelic expression in single cells. ASB of TFs to distal enhancer regions and transcriptional start sites explain 74~86% of the variance for transcriptional initiation rate and burst frequency, whereas they only explain ~25% for burst size. Moreover, for 108,156 eQTL identified in LCLs, we found a significant correlation between the effect size of eQTL and the fold change of promoter kinetics between two alleles, specifically burst frequency(coefficient=0.283,p value=2.073e-09). Previous studies in mammalian cells also show a tight regulation of burst frequency rather than burst size. The transcriptome-wide analysis shows that the ASB of YY1 and EED are associated with the change of promoter kinetics in LCLs(FDR<0.1). Moreover, for a single chromosome, chromatin accessibility together with ASB of TFs shapes the promoter kinetics. A suggestive eQTL, rs4148869, has a C allele concurrent with 5 TFs showing ASB, unmethylated CpG islands, allelic ATAC-seq peaks, and larger promoter kinetics of HLA-DQB1. A GWAS locus rs381218 was potentially associated with autoimmune traits due to ASB of TBP and allelic open chromatin. In summary, the differences in genetic and epigenetic factors between two chromosomes regulate the promoter kinetics, leading to allelic specific expression, which possibly helps explain the mechanism of eQTL and haploinsufficiency.

S31. How do we express ourselves? Examining promoter and enhancer biology

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 195. Massively parallel reporter assay reveals promoter position-dependent and tissue-specific effects in islet TSSs

Authors:

M. Bose, A. Tovar, A. Varshney, Y. Kyono, J. Kitzman, S. Parker; Univ. of Michigan, Ann Arbor, MI

Abstract:

Massively parallel reporter assays (MPRA) are a high-throughput method to assess activity of candidate cis-regulatory sequences, and can be used to detect allelic differences at disease-associated variants. Previously published MPRA studies have screened functional SNPs associated with various traits and conditions, including type 2 diabetes. Here, we used an MPRA library to quantify activity of 1,305 pancreatic islet-derived TSSs generated from CAP analysis of gene expression (CAGE) profiling. We cloned oligos upstream or downstream of a reporter gene along with either the human insulin (INS) promoter or a synthetic housekeeping promoter (SCP1). We used generalized linear models (GLM) to predict position-specific oligo activity from tissue-specific chromatin state regulatory annotations. Overlap with islet ATAC-seq peaks ($\beta = 9.20$, $p < 0.01$) and positive strandedness ($\beta = 11.91$, $p < 0.001$) both predicted significantly higher activity for oligos cloned in the upstream position. We also used Lasso regression to predict position-specific oligo activity based on enrichment of transcription factor motifs (TFMs). Oligos that displayed preferential activity upstream of the reporter gene were enriched for TFMs such as FOXJ2, and BHLHE22, and E2F. These significant differences were not observed for oligos that had higher activity when cloned downstream of the promoter and reporter gene. Together, these results support use of MPRA strategies that account for positional and promoter context-dependent factors when assaying candidate regulatory elements in pursuit of understanding complex genetic diseases.

S31. How do we express ourselves? Examining promoter and enhancer biology

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 196. Adaptive sequence divergence forged new neurodevelopmental enhancers in humans

Authors:

R. Mangan, F. C. Alsina, F. Mosti, J. Sotelo-Fonseca, D. A. Snellings, E. Au, J. Carvalho, L. Sathyan, G. D. Johnson, T. E. Reddy, D. L. Silver, C. B. Lowe; Duke Univ., Durham, NC

Abstract:

The evolutionary origins of distinctly human phenotypes, including bipedalism and brain expansion, lie in the genetic differences between great apes and modern humans. Previous searches for the genetic underpinnings of uniquely human traits were limited to highly conserved genomic regions to identify adaptive innovations of existing functional elements. However, genomic regions of past conservation exclude regions with recurrent modifications and novel functional elements descended from previously neutral regions. Here we demonstrate that the fastest-evolving regions of the human genome, which we term Human Ancestor Quickly Evolving Regions (HAQERs), rapidly diverged in an episodic burst of directional positive selection prior to the Human-Neanderthal split before transitioning to constraint among modern humans. HAQERs are depleted from coding regions and strongly overrepresented in human-specific gene regulatory elements and in bivalent chromatin. We developed in vivo single-cell STARR-seq as a multiplex enhancer assay in developing tissues to reveal that rapid HAQER divergence generated new hominin-specific enhancers in the developing brain. HAQERs are associated with key neurodevelopmental genes and enriched for disease-linked variation. We propose that a lack of pleiotropic constraints and elevated mutation rates poised HAQERs for rapid adaptation and subsequent susceptibility to disease.

S31. How do we express ourselves? Examining promoter and enhancer biology

Location: Conv Ctr/Room 515/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 197. A global high-density chromatin interaction network reveals functional long-range and trans-chromosomal relationships

Authors:

R. Iohia[1], N. Fox[1], J. Gillis[2]; [1]Cold Spring Harbor Lab., Cold Spring Harbor, NY, [2]Univ. of Toronto, Toronto, ON, Canada

Abstract:

Chromatin contacts are essential for gene-expression regulation, however, obtaining a high-resolution genome-wide chromatin contact map is still prohibitively expensive owing to large genome sizes and the quadratic scale of pairwise data. Chromosome conformation capture (3C) based methods such as Hi-C have been extensively used to obtain chromatin contacts. However, since the sparsity of these maps increases with an increase in genomic distance between contacts, long-range or trans chromatin contacts are especially challenging to sample.
Here, we created a high density reference genome-wide chromatin contact map using a meta-analytic approach. We integrate 3600 Human, 6700 Mouse, and 500 Fly 3C experiments to create species-specific meta-3C contact maps with 304 billion, 193 billion, and 19 billion contacts in respective species. We validate that meta-3C are uniquely powered to capture functional chromatin contacts in both cis and trans. Unlike individual experiments, meta-3C gene contacts predict gene coexpression for long-range and trans chromatin contacts. Similarly, for long-range cis-regulatory interactions, meta-3C contacts outperform all individual experiments, providing an improvement over the conventionally used linear genomic distance-based association. Assessing between species, we find patterns of chromatin contacts conservation in both cis and trans and strong associations with coexpression even in species for which 3C data is lacking.
We have generated an integrated chromatin interaction network which complements a large number of methodological and analytic approaches focused on improved specificity or interpretation. This high-depth “super-experiment” is surprisingly powerful in capturing long- range functional relationships of chromatin interactions, which are now able to predict coexpression, expression quantitative trait loci (eQTL), and cross-species relationships.

S32. Influence of germline variants on somatic genomic features in cancer risk and progression

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 200. Germline cancer gene expression quantitative trait loci influence local and global tumor mutations

Authors:

Y. Liu[1], A. Gusev[2], P. Kraft[1]; [1]Harvard Sch. of Publ. Hlth., Boston, MA, [2]Dana-Farber Cancer Inst. and Harvard Med. Sch., Boston, MA

Abstract:

Somatic mutations drive cancer development and are relevant to patients' response to treatment. Emerging evidence show that variations in somatic genome are influenced by germline genetic background. However, the mechanisms underlying these germline-somatic associations remain largely obscure. We hypothesize that germline variants can influence somatic mutations in a nearby cancer gene ("local impact") or a set of recurrently mutated cancer genes across the genome ("global impact") through their regulatory effect on gene expression. Here, by integrating tumor targeted sequencing data from 12,413 patients across 11 cancer types in the Dana-Farber Profile cohort with germline cancer gene expression quantitative trait loci (eQTL) data from the Genotype-Tissue Expression Project, we identified novel associations between cancer gene eQTL and tumor mutations. For local impact, we found 11q22.3 variants that upregulate *ATM* expression which are also associated with a decreased risk of having somatic *ATM* mutations across 8 cancer types ($P = 3.43 \times 10^{-5}$). For global impact, we identified *GLI2*, *WRN*, and *CBFB* eQTL that are associated with tumor mutational burden of cancer genes in ovarian cancer, glioma, and esophagogastric carcinoma, respectively, with $P < 3.45 \times 10^{-6}$. An *EPHA5* eQTL was associated with tumor mutation count (TMC) of cancer genes in colorectal cancer. eQTL associated with expression of *APC*, *WRN*, *GLI1*, *FANCA*, and *TP53* were associated with TMC in endometrial cancer ($P < 1.73 \times 10^{-5}$). Our findings provide evidence for the germline-somatic associations mediated through the expression of specific cancer genes and open new avenues for research on the underlying biological processes, especially those related to immunotherapy responses.

S32. Influence of germline variants on somatic genomic features in cancer risk and progression

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 201. A comprehensive analysis of clinical and polygenic germline influences on somatic mutational burden with implications for survival

Authors:

K. Taraszka[1], S. Groha[2], D. King[3], R. Tell[3], K. White[3], N. Zaitlen[1], A. Gusev[2]; [1]Univ. of California, Los Angeles, CA, [2]Dana-Farber Cancer Inst. and Harvard Med. Sch., Boston, MA, [3]Tempus, Chicago, IL

Abstract:

Cancer is caused by cancer-gene dysregulation which is often induced by somatic alterations. Previous work has linked clinical features and polygenic scores (PGS) to somatic mutation profiles and has identified total mutational burden (TMB) as a biomarker for immunotherapy. Unfortunately, the opportunity to comprehensively study somatic mutational burden has been limited by the sample size and scope of previous studies. Here, we present the largest germline-somatic study to date with >23,000 patients of European ancestry. The discovery cohort consists of 17 cancers with tumors collected during routine care (N=13,131 pan-cancer) while the replication cohort contains 14 cancers from a direct to consumer setting which offers tumor and normal-match sequencing (N=10,294). All samples leverage off-target imputation to generate germline calls. We consider two measures of somatic mutational burden: 1) TMB which is the enumeration of somatic single nucleotide variants; 2) CNB which is the total number of somatic copy number variants. In addition to identifying features associated with TMB/CNB, we conduct follow-up survival analyses on the significantly associated features. We identified 23 novel discoveries and confirm 11 previously established associations between TMB/CNB and the clinical features: age, sex, and metastatic status ($p<3.85x10^{-3}$). We tested 21/34 events for replication and observed 12 associations replicate ($p<0.05$) with 6 remaining after multiple testing correction. To determine if polygenic risk for phenotypes is linked with TMB/CNB we tested the PGS for 14 traits. We observed 16 discoveries ($p<3.57x10^{-3}$), 15 of which are novel: 12 per-cancer and 4 pan-cancer. We tested 11/16 for replication where 3/7 per-cancer events replicate and 1/4 pan-cancer associations replicate. We next explored the effect of fine-scale genetic ancestry on TMB/CNB along the Northwest-Southeast (NW-SE) cline and Ashkenazi Jewish - non Jewish (AJ-nonAJ) cline. We found in both cohorts increased NW ancestry and separately increased AJ ancestry are associated with lower TMB across cancers. Finally, we explored whether PGS or ancestry were linked with survival. We found AJ ancestry is protective in two cancers and SE ancestry is protective in one, all three effects remain after conditioning on TMB. We also identified two protective interaction effects on survival. In melanoma, the PGS for lung cancer modifies the effect of CNB while NW-SE ancestry modifies the effect of TMB in prostate cancer. While much of the accumulation of somatic alterations is a stochastic process, our work indicates there are host characteristics (e.g. age, PGS, genetic ancestry) that shape it.

S32. Influence of germline variants on somatic genomic features in cancer risk and progression

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 202. The impact of common germline risk on somatic alterations and clinical features across cancers

Authors:

S. Namba[1], Y. Saito[2,3], Y. Kogure[2], T. Masuda[1,4], Y. Okada[1,5,6,7], K. Kataoka[2,3]; [1]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]Natl. Cancer Ctr. Res. Inst., Tokyo, Japan, [3]Keio Univ. Sch. of Med., Tokyo, Japan, [4]StemRIM Inst. of Regeneration-Inducing Med., Suita, Japan, [5]Graduate Sch. of Med., the Univ. of Tokyo, Tokyo, Japan, [6]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [7]Immunology Frontier Res. Ctr. (WPI-IFReC), Osaka Univ., Suita, Japan

Abstract:

Cancer is a genomic disease driven by the accumulation of somatic alterations, but germline variants also contribute to the process of carcinogenesis. Recent genome-wide association studies (GWASs) have identified hundreds of common germline risk variants, and the aggregation of genome-wide risk variants, known as polygenic risk score (PRS), can explain genetic liability to cancers. However, there are only sporadic reports on PRS associations with certain driver mutations and cancer subtypes. A comprehensive evaluation of polygenic germline-somatic associations is thus needed. We constructed PRSs from 14 GWASs (median n = 64,905) for 12 cancer types using multiple methods (clumping and thresholding, LDpred2, lassosum, and PRScs). We evaluated the PRSs in the UK Biobank (n = 335,048) and observed that seven PRSs (for seven cancer types) had high predictive performance (i.e., $R^2 > 0.01$ and adjusted P-value < 0.05). By jointly modeling PRSs and rare germline pathogenic variants, we examined the PRS associations with a wide range of somatic alterations and clinical features in The Cancer Genome Atlas (TCGA) (n = 2,921) to provide a comprehensive portrait of germline-somatic associations. Meta-analyses across cancer types revealed that higher PRS was associated with earlier cancer onset (P = 0.001). Higher PRS was also associated with a lower burden of somatic alterations, including total mutations, chromosome/arm somatic copy number alterations (SCNAs), and focal SCNAs (P = 0.032, 0.008, and 0.040, respectively). As increased genomic instability is characteristic of later stages of carcinogenesis, these results suggest that common germline risk enables early tumor development before many mutations and SCNAs accumulate. The associations between PRS and somatic alterations showed no apparent heterogeneity across cancer types, in contrast to rare germline pathogenic variants associated with somatic alterations in cancer type-specific manners. The associations with PRS were not significant for the number of driver mutations or mutations in individual driver genes, suggesting that common germline risk does not necessarily affect the necessity of driver mutations during cancer development. To independently validate the germline-somatic associations, we analyzed additional three prostate cancer cohorts (n = 32, 40, and 116) and confirmed the associations between PRS and somatic alterations to that from TCGA. Our work provides the best available evidence that the overall effects of PRS on somatic alterations were maintained across cancers, different from rare pathogenic variants.

S32. Influence of germline variants on somatic genomic features in cancer risk and progression

Location: Conv Ctr/Petree C/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 203. Genomic analysis of skin cancers from xeroderma pigmentosum subgroups revealed new mechanisms of UV mutagenesis

Authors:

S. Nikolaev[1], A. Yurchenko[1], T. Lajus[2], F. Rajabi[1], P. Kannouche[1], A. Sarasin[1], C. Menck[3], H. Fassihi[4], C. Nishigori[5], A. Lehmann[6]; [1]Inst. Gustave Roussy, Villejuif, France, [2]Departamento de Biologia Celular e Genética, Univ.e Federal do Rio Grande do Norte, Natal, Brazil, [3]Inst. of BioMed. Sci., Sao Paulo, Brazil, [4]St John's Inst. of Dermatology, London, United Kingdom, [5]Kobe Univ. Graduate Sch. of Med., Kobe, Japan, [6]Univ. of Sussex, Brighton, United Kingdom

Abstract:

Rare autosomal syndrome Xeroderma Pigmentosum (XP) is characterised by 1000 times increased risk of skin cancer due to the impaired Nucleotide Excision Repair (NER) pathway or translesion synthesis (polymerase eta). We assembled a unique collection of skin tumours (n=39) from five most frequent and cancer-prone XP subgroups (XP-A, XP-C, XP-D, XP-E, XP-V) and performed whole genome sequencing to characterise in detail their genomic mutational landscapes comparing with tumour type-matched sporadic cancers (n=139). The ultramutated tumour phenotype was observed in the samples with impaired GG-NER (XP-E=370 mut/Mb, XP-C=153 mut/Mb) and translesion synthesis (XP-V=330 mut/Mb), while it was significantly lower and comparable with sporadic cancers in the groups with mutations in TFIIH-XPA complex (XP-A=36 mut/Mb, XP-D=48 mut/Mb). XP-C and XP-E samples were characterized by very strong transcriptional bias. XP-V tumors with absent polymerase eta were characterised by distinct mutational signature with 28% of all SNVs represented by G>T mutations which are very rare in sporadic skin cancer. These mutations demonstrated strong transcriptional bias indicating that the lesions are a substrate of NER and are bypassed in error-free manner in sporadic cancers by polymerase eta. Another important finding was a strong bias in incorrect nucleotide insertion in 3' relative to 5' prime of pyrimidine dimer in XP-V group (40-fold) suggesting that polymerase eta is indispensable as inserter polymerase at UV lesions. We then observed that mutation rates in XP groups with complete lack of NER do not depend on replication timing. Strikingly, in XP-D group there was no replication bias, opposingly to the other groups of skin cancer where strong enrichment of UV mutations on lagging strand was observed. XP-D and XP-C groups were also different from the other groups by very high percentage (20%) of double base substitutions. In summary, different XP-groups demonstrated unique UV mutational patterns, with 10-30% of mutations not explained by the COSMIC UV signatures SBS7a-d. This suggests that mutational profiles in skin cancer depend on efficacy of NER and TLS.

S33. Investigating complex genomic regions

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 206. Repeat polymorphisms underlie top genetic risk loci for glaucoma and colorectal cancer

Authors:

R. Mukamel[1], R. Handsaker[2], M. Sherman[3], A. Barton[4], M. Hujoel[5], S. McCarroll[1], P-R. Loh[5]; [1]Harvard Med. Sch., Boston, MA, [2]Broad Inst., Cambridge, MA, [3]Massachusetts Inst. of Technology, Cambridge, MA, [4]Harvard Univ., Cambridge, MA, [5]Brigham and Women's Hosp./Harvard Med. Sch., Boston, MA

Abstract:

Many regions in the human genome vary in length among individuals due to variable numbers of tandem repeats (VNTRs). We recently showed that VNTRs overlapping protein-coding exons underlie some of the strongest known genetic associations with diverse phenotypes. Here, we assessed the phenotypic impact of VNTRs genomewide, the vast majority of which lie in noncoding regions. We estimated VNTR allele lengths in *N*=418,136 unrelated UK Biobank participants of European ancestry using statistical phasing and imputation of allele lengths estimated from whole-genome sequencing depth-of-coverage. Among 15,653 autosomal, multiallelic repeat loci (ascertained by analysis of *N*=64 long-read assemblies from HGSVC2), this genotyping strategy typically captured a substantial proportion of allelic variation (median R^2=0.48), with the most variable VNTRs (allele length s.d. >100bp, *N*=4,462 loci) particularly well-genotyped (median R^2=0.79).
Association and fine-mapping analyses of 786 phenotypes identified 4,968 significant VNTR-phenotype associations, 107 of which (involving 58 VNTRs) were assigned a high probability of causality by statistical fine-mapping (PIP>0.5). Noncoding VNTRs appeared to explain some of the strongest known SNP associations with glaucoma and colorectal cancer. At *TMCO1*, the length of an intronic 28bp repeat associated more strongly with glaucoma risk than any SNP or indel in the entire genome (P=3.1 x 10^{-71}), with excess cases among carriers of expanded alleles accounting for ~5% of primary-open angle glaucoma cases in UK Biobank. The association appeared to be driven by a series of expanded alleles (5-11 repeats, combined AF=12%), with the longest alleles (AF=1%) conferring 1.79-fold increased glaucoma risk (95% CI=1.51-2.12). Downstream of *EIF3H*, the length of a 27bp repeat (2-6 repeats per allele) associated with increasing risk of colon polyps (P=8.2 x 10^{-35}) and colorectal cancer (P=3.5 x 10^{-25}; OR=1.4 for longest vs. shortest alleles). The explanatory power of this locus, which again ranked first among all colorectal cancer loci genome-wide, was underestimated by ~50% in previous association studies that considered only SNPs in partial LD with the VNTR. Additionally, at *CUL4A*, expansion of a highly polymorphic intronic ~32bp repeat (allele length range >2.5kb) associated with reduced mean corpuscular hemoglobin (P=7.8 x 10^{-41}) and other hematological traits, apparently driven by repeat-mediated cryptic splicing of *CUL4A* resulting in the loss of 15 terminal exons (P=9.3 x 10^{-80} in GTEx RNA-seq data). These results indicate a substantial and previously underappreciated role of noncoding VNTRs in shaping human phenotypes.

S33. Investigating complex genomic regions

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 207. Extensive mosaicism by somatic L1 retrotransposition in normal human cells

Authors:

Y. Ju; Korea Advanced Inst. of Sci. and Technology, Daejeon, Korea, Republic of

Abstract:

Over the course of an individual's lifetime, genomic alterations accumulate in somatic cells. However, the mutational landscape by retrotranspositions of long interspersed nuclear element-1 (**L1**), a widespread mobile element in the human genome, is poorly understood in normal cells. Here, we explored the whole-genome sequences of 892 single-cell clones established from various tissues collected from 28 individuals. Remarkably, 88% of colorectal epithelial cells acquired somatic L1 retrotranspositions (**soL1Rs**), carrying ~3 events per cell on average with substantial intra- and inter-individual variances, which was accelerated at least 10-fold during tumourigenesis. Breakpoints of soL1Rs suggested that a few variant mechanisms can be involved in the L1 retrotransposition processes. Fingerprinting of donor L1s using source-specific unique sequences revealed 34 hot L1s, 44% of which were newly discovered in this study, and many ultra-rare hot L1s in the human population showed higher retrotransposition potential in somatic lineages than common sources. Multi-dimensional analysis of soL1Rs with early embryonic developmental relationships, genome-wide methylation, and gene expression profiles of the clones demonstrated that (1) soL1Rs occur from early embryogenesis at a substantial rate, (2) epigenetic activation of hot L1s is stochastically acquired during the wave of early global epigenomic reprogramming, rather than by the sporadic loss-of-methylation at the late stage, and (3) most L1 transcripts in the cytoplasm do not generate soL1Rs in somatic lineages. In summary, this study provides insights into the retrotransposition dynamics of L1s in the human genome and the resultant somatic mosaicism in normal human cells.

S33. Investigating complex genomic regions

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 208. Evaluation of long reads across challenging medically relevant genes and their implications for *All of Us*

Authors:

M. Mahmoud[1], Y. Huang[2], K. Garimella[2], P. Audano[3], W. Wan[2], N. Prasad[4], R. Handsaker[2], S. Hall[5], A. Pionzio[4], M. Talkowski[6], E. Eichler[7], S. Levy[5], F. Sedlazeck[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Broad Inst., Cambridge, MA, [3]The Jackson Lab., Farmington, CT, [4]Discovery Life Sci., Huntsville, AL, [5]HudsonAlpha Inst. for Biotechnology, Huntsville, AL, [6]Massachusetts Gen. Hosp., Boston, MA, [7]Univ. of Washington, Seattle, WA

Abstract:

The overall heritability explained by genetic studies to date is low for most complex diseases. Several common variant studies have suggested associations with loci in proximity to complex genomic regions that, in aggregate, harbor ~400 medically relevant genes (e.g., *SMN1, RHCE, NCF1, LPA, TPO, TNNT1,* and *HLA*), many of which are refractory to alignment using short-read genome sequencing. The emergence of long-read technologies holds great promise for the detection and phasing of short (SNVs and indels) and structural variants (SVs) to disentangle the complex interactions of these mutations. Unfortunately, the error rate and high costs limit availability and application in clinical sequencing. The *All of Us* (AoU) long-read research project aims to broaden our understanding of the sequence diversity in complex genomic regions that potentially have direct implications on variation discovery in medically relevant genes, which paves the way for individualized prevention, treatment, and care. We previously identified 386 genes largely inaccessible using short reads, which are also found in genetic panels and highlighted in multiple studies, citing their medical importance. Moreover, mutations in these genes cause different diseases such as cancer, spinal muscular atrophy, and Rh deficiency syndrome. Using these genes plus 5,027 medically relevant genes reported across studies, we investigate the utility and limitations of long-read sequencing in analyzing HapMap samples plus four conditions of two samples using different tissue and extraction methods with Illumina, PacBio HiFi, and Oxford Nanopore Technologies (ONT). In the present study, we develop a cloud workflow to analyze long-read data utilizing state-of-the-art tools for SNVs, indels, and SVs. Likewise, we identify the best combination of tools and technologies to leverage SNVs and indels precision, recall, and F-score (0.999, 0.998, and 0.999). Furthermore, we highlight the utility of long-read sequencing in calling SVs in complex medically relevant genes, which enabled us to call SVs with a 0.93 F-score, exceeding Illumina 0.45. Moreover, we assess the advantages and drawbacks per technology in covering ClinVar and GIAB truth set variant identification. To conclude, given similar coverage levels, HiFi outperforms other technologies. Nevertheless, we identify several genes where ONT (*SMN1*, *KSCNE1*, *MUC1*, and *TERT*) or Illumina (*KMT2C*) show improved variant identification compared to HiFi. My presentation details the benefits and risks of long-read technology within AoU samples and contrasts the results with achievements on the Illumina side with the recommended AoU analysis pipeline.

S33. Investigating complex genomic regions

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 209. A global reference for human genetic variation at tandem repeats.

Authors:

H. Ziaei Jam[1], N. Mousavi[1], N. Ma[1], R. DeVito[1], B. Huang[1], M. Maksimov[1], E. Dolzhenko[2], Y. Qiu[2], M. Bakhtiari[1], J. Park[1], V. Bafna[1], M. Gymrek[1]; [1]Univ. of California San Diego, La Jolla, CA, [2]Illumina Inc., San Diego, CA

Abstract:

Tandem repeats (TRs) are highly variable elements of the human genome consisting of repeated sequences of motifs of 1 or more nucleotides. There are more than 1 million unique TR loci in the human genome. Despite their abundance, TRs have proven to be difficult to study at a large scale due to technical challenges they present in sequencing and genotyping. A variety of tools have been developed to genotype TRs using short-read sequencing data, including HipSTR, GangSTR, adVNTR, and ExpansionHunter. These tools have been applied to study TR variation at population-scale across multiple cohorts. However, existing methods are designed to call largely non-overlapping types of TRs, which only together capture the full spectrum of TR variation. Further, existing TR callsets from these methods have been largely biased toward individuals of European descent. Here, we develop EnsembleTR, a graph-based method that integrates genotypes from the multiple distinct TR callers above to report consensus genotypes for each TR. We applied EnsembleTR to high-coverage whole-genome sequencing for 2,504 unrelated samples from the 1000 Genomes Project to generate a deep reference panel of TR variation across worldwide populations at ~1.8 million TRs. We performed fragment analysis to experimentally validate our calls at 76 TRs across 48 samples, and found >94% concordance with EnsembleTR.
Using the improved data from EnsembleTR, we studied patterns of TR variation across populations. Our catalog recapitulates known patterns, such as higher variation rates in African populations and the relationship between repeat length and heterozygosity rates. We also uncover novel findings, including (1) context sequence features in the immediate flanking regions of repeats influencing TR stability, (2) population-specific repeat expansions, including an experimentally validated CAG repeat in CA10 that is specifically expanded in African populations, and (3) TRs showing strong signals of negative selection in one or more populations. Finally, we generate a phased reference panel that allows accurate imputation of TR variants from SNP data. Our results and analysis provide a genome-wide TR catalog as well as population-level TR characterizations that will open the path for further studies.

S34. New methods for revealing repeats

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 212. RExPRT: a machine learning tool to rank repeat expansions by pathogenicity

Authors:

S. Fazal, M. Danzi, S. Zuchner, V. Aguiar-Pulido; Univ. of Miami, Miami, FL

Abstract:

Expansions of short tandem repeats (TRs) are responsible for ~50 known diseases, most of which primarily affect the nervous system. We speculate these represent only a fraction of the pathogenic repeat expansions that exist and that they may be responsible for explaining a proportion of the diagnostic gap in rare monogenic diseases. The identification of TRs from WGS has become more accessible in recent years with the advancement of tools such as ExpansionHunter and GangSTR. However, the challenge of distinguishing novel pathogenic expansions from those that are benign remains. Pioneering these efforts, we present RExPRT, a machine learning-based Repeat EXpansion Pathogenicity Ranking Tool.
RExPRT was trained on 40 known disease-associated TRs and 755 commonly expanded benign TRs. Predictive features for RExPRT are annotations of the surrounding genetic architecture that were selected based on their enrichment in pathogenic loci compared to other repeats. Leave-one-out cross validation results demonstrated that the support vector machine (SVM) model led to the highest accuracy in classification, with a precision of 100% and a recall of 70%. With the ensemble of SVM and the gradient boosted decision tree model, a total of 30 disease-associated TRs (75%) were recalled. We also tested RExPRT's ability to rank rare expansions from 82 positive controls for various repeat expansion disorders. Of samples with pathogenic TRs identified by the variant callers, RExPRT ranked the disease-causing TR 1st among other presumably benign TR expansions in 67% of samples. Only three out of 182 benign TRs were falsely classified as pathogenic. The extremely low false positive rate and excellent prioritization of pathogenic expansions by RExPRT demonstrates its immense value in selecting strong candidates to push forward into functional studies, which are time-consuming and costly. RExPRT has established the possibility of using machine learning to prioritize repeat expansion variation. It lays the groundwork for similar tools that can be built in the future, mitigating some of its limitations.

S34. New methods for revealing repeats

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 213. Precise and ultrafast tandem repeat variant detection in massively parallel sequencing reads.

Authors:

X. Wang[1], M. Huang[1], B. Budowle[1,2], J. Ge[1,2]; [1]Univ. of Texas Hlth.Sci. center, Fort Worth, TX, [2]Dept. of Microbiol., Immunology, and Genetics, Univ. of North Texas Hlth.Sci. Ctr., Fort Worth, TX

Abstract:

Calling tandem repeat (TR) variants from DNA sequences is of both theoretical and practical significance. A large number of software tools have been developed for detecting TRs. However, little study has been done to detect TR alleles from long-read sequences, and the effectiveness of detecting TR alleles from whole-genome sequence (WGS) data still needs to be improved. Herein, a novel algorithm is described that determines the boundaries of TR regions, and a software program, TRcaller, has been developed to call TR alleles from both short- and long-read sequences, both whole-genome and targeted sequences generated from multiple sequencing platforms. The results showed that TRcaller can provide substantially higher accuracy in detecting TR alleles with magnitudes faster than the mainstream software tools. 99.4% call accuracy has been achieved for 20 CODIS core STR loci from 289 WGS data samples with 30x coverage of Illumina reads from the 1000 genomes project, which is higher than that from HipSTR (i.e., 93.4%). To reach a 99.9% calling rate, at least 25x, 10x, and 5x average depths were needed for Illumina PE150, Illumina PE250, and PacBio CCS reads, respectively. TRcaller takes less than 2 seconds for calling STRs at CODIS core STR set from WGS sequencing reads up to 300x. TRcaller is able to facilitate scalable, accurate, and ultrafast TR allele calling from large-scale sequence datasets in various applications, such as forensics, medical research, disease diagnosis, clinical testing, evolution, and breeding programs.

S34. New methods for revealing repeats

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 214. Genome-wide analysis of mutations and epigenetic changes of tandem repeat regions in rare genetic disease cases

Authors:

E. Dolzhenko[1], W. J. Rowell[1], H. Brand[2], M. A. Brown[3], K. Chua[1], G. Elgar[3], G. S. Erwin[4], H. Houlden[5], K. Ibanez[3], D. Kasperaviciute[3], Z. Kronenberg[1], N. E. Kurtas[2], A. Rendon[3], R. Sullivan[6], A. Tucci[3], J. Vandrovcova[7], B. Weisburd[2], A. Wenger[1], M. A. Eberle[1]; [1]Pacific BioSci.s of California, Menlo Park, CA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Genomics England, Queen Mary Univ. of London, London, United Kingdom, [4]Stanford Univ., Stanford, CA, [5]Natl. Hosp. for Neurology and Neurosurgery, London, United Kingdom, [6]UCL Queen Square Inst. of Neurology, London, United Kingdom, [7]UCL Queen Square Inst. Of Neurology, London, United Kingdom

Abstract:

DNA mutations and epigenetic changes in tandem repeat (TR) regions have a well-established association with Mendelian disease, cancer, and many complex traits. These regions are only partially characterized by short-read whole-genome sequencing, with typical pipelines evaluating only a small subset of TRs in the human genome. Although long-read sequencing is able to span TRs, comprehensive analysis has been limited by informatics pipelines that classify TR variants as simple indels or structural variants. To enable accurate and consistent interpretation of TR variants, we have developed a suite of tandem repeat genotyping (TRGT) methods for analysis and visualization of TR regions from HiFi reads. TRGT provides haplotype-resolved estimates of lengths, mosaicism, CpG methylation, and sequence composition for each analyzed repeat.

To assess TRGT accuracy, we analyzed 500,000 polymorphic tandem repeats across healthy trio controls. Over 96% of repeat sites were Mendelian consistent and allowing one repeat unit mismatches increased the consistency rate to over 99.5%. We also analyzed samples with known pathogenic repeat expansions in HTT and FMR1 repeats. The genotypes and mosaicism estimates produced by TRGT were consistent with what is known about these samples. For example, for NA07537 our method detected a hypomethylated short allele with two AGG interruptions spanning 87 base pairs and a hypermethylated mosaic FMR1 expansion with sizes ranging from 840 to 1200 base pairs.

We next used TRGT to analyze 56 known disease-causing repeats in HiFi sequencing data from 62 participants from the 100,000 Genomes Project Rare Disease Arm. We identified likely pathogenic repeat expansions in 14.5% (9/62) of the samples including: a hypermethylated mosaic FMR1 repeat expansion spanning between 386 and 519 CGG motifs; an ATXN8 expansion spanning 577 CTG motifs; seven biallelic RFC1 repeat expansions with 186 to 1647 AAGGG motifs. Five of these expansions have been validated and the rest are in the process of being validated. Notably, none of these expanded repeats can be fully characterized with short reads.

In summary, TRGT can determine the most important features of pathogenic repeats from long-read HiFi whole-genome sequencing data including accurate sizing of repeats spanning from 10s to 1000s of base pairs, quantification of methylation patterns, and detection of changes to the repeat composition. We will present these results along with the description of a population-scale resource with lengths and methylation states of 500,000+ tandem repeats.

S34. New methods for revealing repeats

Location: Conv Ctr/Room 502/West Building

Session Time: Wednesday, October 26, 2022, 1:45 pm - 2:45 pm

ProgNbr 215. Non-linear effects of short tandem repeats on gene expression

Authors:

M. Lamkin, Y. Li, M. Gymrek; Univ. of California, San Diego, San Diego, CA

Abstract:

Short tandem repeats (STRs), consisting of consecutively repeated units of 1-6bp, are among the most polymorphic regions of the human genome. Recent studies have identified thousands of STRs for which variation in repeat length across individuals is associated with complex traits including gene expression, height, and blood traits. Existing STR-based association studies have almost exclusively considered linear models of the effect of repeat length on phenotypes. However, multiple well-studied STRs are known to exhibit non-linear effects, suggesting linear models may not always be most appropriate. For example, most STRs implicated in expansion disorders such as Huntington's Disease follow a threshold model, in which repeat lengths below a certain threshold have no effect. Further, studies of regulatory effects of STRs in yeast have identified STRs with a bell curve-shaped effect, where repeats at some optimal length give maximal expression, but very long or short repeat lengths reduce expression. Testing for non-linear associations between repeat lengths and phenotypes in diploid organisms is challenging, since the effect of each allele must be modeled separately and the repeat lengths cannot simply be summed together. Here, we develop an association testing framework in which we model the effect of each STR allele as an nth-degree polynomial (typically, we set n=2). We evaluate our regression framework using simulated data and demonstrate that our model is well-powered to detect quadratic effects at highly multi-allelic STRs, whereas as expected non-linear effects are more difficult to detect for bi-allelic STRs.

We applied our framework to identify non-linear effects of STRs on gene expression using 654 samples from the Genotype Tissue Expression (GTEx) Project for which we had previously performed genome-wide STR genotyping of 1.6 million STRs. We identified 761 STR effects on expression of genes in the tibial nerve tissue for which a quadratic association provided the best model fit (ANOVA $p<6.157e-8$). Interestingly, in many of these cases linear association testing did not identify a significant association. Overall, our results suggest linear models used so far are not sufficient to capture more complex associations between STR repeat length and phenotype, and that these associations are also unlikely to be well-tagged by nearby common single nucleotide variants. In future work, we aim to incorporate these non-linear associations into polygenic prediction scores for gene expression and complex traits.

S38. Disorders of bone pathology X-rayed using common and rare disease approaches

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 234. EFEMP1 regulates bone structure and effects fracture risk: A discovery from HRPQCT bone microarchitecture GWAS screening confirmed by CRISPR knockout zebrafish and mice: The Bone Microarchitecture International Consortium (BoMIC)

Authors:

Y-H. Hsu[1], R. Kwon[2], F. Kinyua[3], A. Wong[4], J. Rai[5], J. Hulleman[6], C. Watson[5], P. Okoro[7], M-J. Tsai[8], M. Nethander[9], T. Lu[10], M. Marian Renwick[6], C-T. Liu[11], D. Karasik[12], S. Boyd[13], M. Bouxsein[14], D. Mellström[15], M. Lorentzon[15], E. Sornay-Rendu[16], E. Lespessailles[17], B. Richards[10], E. Orwoll[18], S. Ferrari[19], S. Khosla[20], P. Szulc[16], R. Chapurlat[21], C. Ohlsson[15], D. Kiel[22]; [1]Marcus Inst. for Aging Res., Harvard Med. Sch., Boston, MA, [2]Univ of Washington, Seattle, WA, [3]HSL Marcus Inst. for Aging Res., Harvard Med Sch., Chestnut Hill, MA, [4]Sch. of public Hlth., Univ. of Toronto, Toronto, ON, Canada, [5]Orthopaedics and Sports Med., Univ. of Washington, Seattle, WA, [6]Ophthalmology, Sch. of Pharmacology, UT Southwestern Med. Ctr., Dallas, TX, [7]Hebrew SeniorLife - Harvard Med. Sch., Roslindale, MA, [8]Marcus Inst. for Aging research, Newton Highlands, MA, [9]Bioinformatics Core Facility, Sahlgrenska Academy, Univ. of Gothenburg, Gothenburg, Sweden, [10]McGill Univ., Montreal, QC, Canada, [11]Boston Univ. SPH, Boston, MA, [12]HSL Marcus Inst. for Aging Res., Harvard Med Sch., Boston, MA, [13]Dept. of Radiology, Cumming Sch. of Med., Univ. of Calgary, Calgary, AB, Canada, [14]Beth Israel Deaconess Med. Ctr., Harvard Med. Sch., Boston, MA, [15]Dept. of Internal Med. and Clinical Nutrition, Inst. of Med., Sahlgrenska Osteoporosis Ctr., at the Sahlgrenska Academy, Univ. of Gothenburg, Gothenburg, Sweden, [16]INSERM and Univ. of Lyon, Lyon, France, [17]Université Orleans, Orleans, France, [18]Oregon Hlth.& Sci. Univ, Portland, OR, [19]Dept. of Med., Univ. of Geneva, Geneva, Swaziland, [20]Mayo Clinic Coll. of Med, Rochester, MN, [21]Claude Bernard Univ. Lyon, Villeurbanne, France, [22]Hebrew SeniorLife, Boston, MA

Abstract:

Both bone mineral density (aBMD) and bone structure affect skeletal integrity. Although bone microarchitecture is associated with osteoporotic fracture risk, no GWAS were conducted on bone microarchitecture. We hypothesized that genetic studies of bone microarchitecture assessed by high resolution peripheral quantitative computed tomography (HR-pQCT) may reveal novel genes that contribute to skeletal integrity. The largest GWAS meta-analysis was conducted in 5,692 adult Caucasians with HR-pQCT measurements (FHS, Mayo Clinic, GeRiCo, OFELY, STRAMBO, MrOS Sweden and GOOD Studies) and replicated in 2,636 independent subjects (MrOS US, CaMOS and QUALYOR studies). Of the 15 replicated GWAS loci ($p < 2.5 \times 10^{-9}$), 6 were not reported by previous BMD GWAS. Most of these loci were specifically associated with specific bone microarchitecture phenotypes including cross-sectional area (TtAr), failure load (µFEA), cortical bone thickness (CtTh), trabecular separation (TbSp), and/or number (TbN). Among novel GWAS loci, predicted functional non-coding SNP rs3791679 in the last intron of *EFEMP1* was associated with lower TtAr (MAF=23%, $p=5\times10^{-10}$) and increased fracture risk (OR=1.2 per allele, 95%CI=1.1-1.4). SNP rs3791679 is in a human osteoblast-specific enhancer (ATAC-seq & Chip-seq in human primary osteoblasts) and predicted to affect binding affinity of TFs POU3F2 and STAT3. In 3D chromatin interactions (HI-C seq in human primary osteoblasts), the active enhancer harboring rs3791679 physically interacted with the promoter of *EFEMP1* gene. Colocalization of gene expression (of human hip bone biopsies) and protein (from plasma proteomics) QTLs also suggested rs3791679 affected EFEMP1 gene/protein expression. EFEMP1 was expressed in human primary osteoblasts (RNA-seq) and in 200 hip bone biopsies (proteomics via MS). CRISPR-edited male & female zebrafishes in efemp1 exhibited significantly increased tissue mineral density at the neural arch (20%↑), centrum (13%↑) and haemal (14%↑) arch; and haemal (11%↑) and neural (10%↑) arch thickness compared to controls. Efemp1 KO mice (male, 3.0-3.6 mo) had significantly denser CtTh (16%↑), higher cortical area (1.9x↑), higher CtBMD (10%↑) and lower TtAr (9%↓) in the tibial diaphysis compared to wildtype mice. Mutations in *EFEMP1* cause heritable marfanoid syndrome, a connective tissue disorder characterized by long limbs and scoliosis. Our finding suggests that studying refined bone microarchitecture phenotypes may reveal novel genetic loci not found using aBMD. Human, zebrafish and mice results suggested common genetic variants act through *EFEMP1* to influence bone structure and fracture risk.

S38. Disorders of bone pathology X-rayed using common and rare disease approaches

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 235. Converging evidence from rare and common variants implicates target genes for osteoporosis

Authors:

S. Zhou[1], K. Sosina[2], J. Bovijn[3], L. Laurent[4], V. Sharma[4], P. Akbari[5], V. Forgetta[4], L. Jiang[4], J. Kosmicki[2], N. Banerjee[6], J. Morris[7], E. Oerton[8], M. Jones[6], M. LeBlanc[9], Regeneron Genetics Center, V. Idone[6], J. Overton[6], J. Reid[6], M. Cantor[6], G. Abecasis[6], D. Goltzman[1], C. Greenwood[10], C. Langenberg[11], A. Baras[6], A. Economides[12], M. Ferreira[6], S. Hatsell[6], B. Richards[1], L. Lotta[6]; [1]McGill Univ., Montreal, QC, Canada, [2]Regeneron Pharmaceuticals, White Plains, NY, [3]Regeneron Genetics Ctr., Oxford, United Kingdom, [4]Lady Davis Inst., Montreal, QC, Canada, [5]Univ. of Cambridge, Cambridge, United Kingdom, [6]Regeneron Genetics Ctr., Tarrytown, NY, [7]New York Genome Ctr., New York, NY, [8]MRC Epidemiology Unit, Univ. of Cambridge, Cambridge, QC, Canada, [9]Regeneron, Bronx, NY, [10]Jewish Gen. Hosp., Montreal, QC, Canada, [11]MRC Epidemiology Unit, Cambridge, United Kingdom, [12]Regeneron Pharmaceuticals, Tarrytown, NY

Abstract:

Common genetic determinants for many human diseases have been described through GWAS, but few studies have leveraged the combined evidence of rare coding variants ascertained by exome sequencing at scale and GWAS of imputed common alleles. We hypothesized that such an approach might identify effector genes and therapeutic targets for osteoporosis, which is clinically defined by reduced bone mineral density. We undertook a large-scale multi-ancestry exome-wide association study for estimated bone mineral density (eBMD) in 278,807 European ancestry individuals and 13,125 individuals of African, East-Asian, or South Asian ancestry from UK Biobank, representing the largest analysis of this trait to date. The burden of rare coding alleles in 19 genes was associated with eBMD at exome-wide statistical significance ($p<3.6x10^{-7}$). These genes were highly enriched for a set of expert-curated Mendelian osteoporosis genes and existing drug targets for osteoporosis (65-fold; $p=2.5x10^{-5}$). By applying a common variant to effector-gene mapping approach (Effector Index (Ei)) to fine-mapped eBMD GWAS signals, we found that exome-wide significant genes had a 2.4-fold higher Ei score than other genes. Further, exome-wide significant genes had 96-fold increased odds of being the top ranked Ei gene at a given locus ($p=1.83x10^{-10}$). Leveraging exome-sequencing, common variant effector-gene mapping, Mendelian randomization and colocalization of 863 circulating proteins, we prioritized *CD109* as a gene for which loss-of-function is associated with higher bone density in humans. CRISPR-Cas9 edits of *CD109* in SaOS-2 osteoblast cell lines demonstrated that decreased CD109 protein increased mineralization by up to 2.4-fold ($p=1.8x10^{-7}$). This study demonstrates that the convergence of exome sequencing, common variant associations, proteomics and CRISPR can pinpoint genes for bone density and highlight novel biology to guide therapeutic development.

S38. Disorders of bone pathology X-rayed using common and rare disease approaches

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 236. COL11A2 as a candidate gene for vertebral malformations and scoliosis

Authors:

P. Giampietro[1], **N. Sobreira**[2], D. Rebello[3], R. Steiner[4], C. Gurnett[5], A. Cornier[6], C. Raggio[7], E. Wohler[2], N. Wu[8], R. Steiner[4], A. Aguilera[1], B. Ciruna[9]; [1]Univ. of Illinois-Chicago, Chicago, IL, [2]Johns Hopkins Univ., Baltimore, MD, [3]Univ. of Toronto, Toronto, ON, Canada, [4]Univ. of Wisconsin Sch. of Med., Marshfield, WI, [5]Washington Univ. in St Louis, Saint Louis, MO, [6]San Jorge Children s Hosp./ Ponce Hlth.Sci. Univ., San Juan, PR, [7]Hosp. for Special Surgery, New York, NY, [8]Peking Union Med. Coll. Hosp., Beijing, China, [9]Hosp. for Sick Children, Toronto, ON, Canada

Abstract:

Human vertebral malformations (VMs) have an incidence of 1/2000 and are associated with health problems including congenital scoliosis (CS), and malformations involving the kidneys, heart and skeleton. The genetic cause for the vast majority of CS is unknown. In a cohort of patients with isolated VM we identified three pathogenic variants in *COL11A2*: R130W, R1407L, R1413H. Interestingly two patients with R130W pathogenic variant had different phenotypes, one having cervical VM and the other have a T9 hemivertebrae. Additionally 24 *COL11A2* variants in CS patients, 7 variants were identified in idiopathic scoliosis (IS) patients and 3 variants observed in individuals with no known scoliosis were identified in the DISCO database (Deciphering Disorders Involving Scoliosis and COmorbidities) which consists of about 800 CS and 400 IS patients, providing evidence for genetic similarities in these conditions. To investigate the loss of function (LOF) a zebrafish phenotype for col11a2, CRISPR/Cas9 was used to create a full deletion of the gene locus, col11a2del. All substitutions observed in our sample cohort are predicted to be damaging to protein function, and R130 and R1407 residues are conserved in zebrafish.To determine the functional consequence of VM-associated variants, we assayed their ability to suppress col11a2del LOF phenotypes following transgenic expression within mutant intervertebral discs. Injection of CRISPR/Cas9 targeting col11a2 produced mosaic mutant animals with a severe vertebral fusion defect. Stable col11a2del/del mutant zebrafish exhibit VMs in caudal vertebrae of the spine. These fusion defects form due to mineralization across the intervertebral discs of the spine. The col11a2del/del animals exhibited a severe phenotype. The LOF phenotype in col11a2 zebrafish can be rescued through transgenic expression of wildtype col11a2 in intervertebral discs of the spine. However, expression of col11a2 R1542L (homologous to human R1407L) failed to rescue VMs in these animals, highlighting an essential role for COL11A2 in vertebral development. We demonstrate that loss of Col11a2 protein function results in the development of vertebral fusions. The mechanism of fusion formation involves mineralization of cartilaginous intervertebral discs, as opposed to defective somite segmentation. Through the establishment of this clinically relevant VM zebrafish model, we demonstrate that loss of Col11a2 protein function results in the development of vertebral fusions.

S38. Disorders of bone pathology X-rayed using common and rare disease approaches

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 237. PTBP1 cytoplasmic retention is associated with osteochondrodysplasia and variable neurodevelopmental anomalies

Authors:

A. Masson[1], J. Paccaud[1], E. Colin[2], O. Mäkitie[3], F. Taylan[4], C. Marcelis[5], D. Koolen[6], S. Robertson[7], L. Bird[8], E. Torti[9], R. Haddad[10], M. Wessels[11], A. Bouman[11], C. Gooch[12], A. Plomp[13], R. LIttlejohn[14], E. Roeder[14], N. Verbeek[15], V. Cormier-Daire[16], M. Mori[17], S. Hickey[17], P. Terhal[15], T. Gardeitchik[18], K. Platzer[19], M. Françoise[20], N. Ekhilevitch[21], Y. Duffourd[22], F. Tran Mau-Them[1,23], C. Philippe[1,23], C. Thauvin-Robinet[1,23,24], L. Faivre[1,25], PTBP1 Research Group, M. Ori[26], Q. Thomas[1,23], A. Vitobello[1,23]; [1]Inserm UMR1231, GAD, Université de Bourgogne, Dijon, France, [2]Univ. Hosp. of Angers, Angers, France, [3]Children's Hosp., Pediatric Res. Ctr., Univ. of Helsinki and Helsinki Univ. Hosp., Helsinki, Finland, [4]Dept. of Clinical Genetics, Karolinska Univ. Lab., Karolinska Univ. Hosp., Stockholm, Sweden, [5]Dept. of Clinical Genetics, Radboud Univ. Med. Ctr., Nijmegen 6525, Netherlands, [6]Dept. of Human Genetics, Radboudumc, Donders Inst. for Brain, Cognition, and Behavior, Nijmegen, Netherlands, [7]Dept. of Women's and Children's Hlth., Dunedin Sch. of Med., Univ. of Otago, Dunedin, New Zealand, [8]Rady Children's Hosp. San Diego, San Diego, CA, [9]GeneDx, Gaithersburg, MD, [10]Div. of Genetics, Dept. of Pediatrics, Medstar Georgetown Univ. Hosp., Washington DC 20007, DC, [11]Dept. of Human Genetics, Erasmus Med. Ctr., Rotterdam, Netherlands, [12]Div. of Genetics and Genomic Med., Dept. of Pediatrics, Washington Univ. Sch. of Med., St Louis, MO, [13]Dept. of Human Genetics, Amsterdam UMC, Univ. of Amsterdam, Amsterdam, Netherlands, [14]Dept. of Pediatrics, Baylor Coll. of Med., San Antonio, TX, [15]Dept. of Genetics, Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [16]Hôpital Necker-Enfants Malades, Paris, France, [17]Div. of Genetic & Genomic Med., Nationwide Children's Hosp., Columbus, OH, [18]Dept. of Human Genetics, Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [19]Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, [20]Ctr. Hosp.ier William Morey, Chalon-sur-Saône, France, [21]The Genetics Inst., Rambam Hlth.Care Campus, Haifa 3109601, Israel, [22]Inserm UMR1231, GAD, Université de Bourgogne, Dijon, France., DIJON, France, [23]Unité Fonctionnelle Innovation en Diagnostic Génomique des Maladies Rares, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France, [24]Ctr. de référence maladies rares « déficiences intellectuelles de causes rares », Ctr. de Génétique, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France, [25]Ctr. de Référence maladies rares « Anomalies du Développement et syndromes malformatifs », Ctr. de Génétique, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France, [26]Dept. of Biology, Univ. of Pisa, Pisa, Italy

Abstract:

Polypyrimidine tract-binding protein 1 (PTBP1) is a ubiquitously expressed heterogeneous nuclear ribonucleoprotein (hnRNP) which acts as a major splicing regulator. Through its partially overlapping nuclear localization signal and export sequence (NLS and NES respectively) and its RNA binding properties, PTBP1 shuttles between nucleus and cytoplasm to play roles in transcript biogenesis and metabolism. Despite the fundamental role as an alternative splicing factor implicated in cell growth, neuronal cell differentiation, and immune cell activation, little is known about its involvement in human disease.
Here, we describe a cohort of 23 individuals with de novo or inherited PTBP1 variants who present with skeletal anomalies (86% of individuals) such as disproportionate short stature, dysplasias affecting the bones of the limbs and other features (e.g. joint hyperlaxity, syndactyly and advanced bone maturation among others). Neurodevelopmental anomalies including evelopmental delay, behavioral problems and intellectual disability were also recurrent and observed in 74% of individuals. Using a combination of molecular genetics and transcriptomics approaches in patient-derived fibroblasts and independent cellular and in vivo models, we demonstrated that pathogenic variants impair NLS/NES function in PTBP1 and cause partial cytoplasmic retention. This results in alteration of mRNA localization and stability. Overall, our data demonstrate that rare, heterozygous variants in PTBP1 affect nucleocytoplasmic distribution and alter biological pathways, leading to skeletal malformations in humans.

S38. Disorders of bone pathology X-rayed using common and rare disease approaches

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 238. Single-cell RNAseq uncovers novel hidden mechanisms in a model of osteogenesis imperfecta

Authors:

J. Zieba, L. Nevarez, D. Wachtell, J. Martin, A. Kot, D. Krakow; Univ. of California, Los Angeles, Los Angeles, CA

Abstract:

Osteogenesis imperfecta (OI), or brittle bone disease, is a rare congenital disorder characterized by bone fragility and increased fracture incidence mainly due to mutations in type I collagen or genes associated with collagen synthesis. Genetic and allelic heterogeneity underlie the phenotypic spectrum of OI and yet all forms commonly feature growth deficiency leading to short stature and early mortality stemming from pulmonary complications the molecular causes for which have not been resolved. Using single-cell RNAseq, we identified novel molecular and cellular mechanisms underlying the cartilage and lung abnormalities observed in our *Col1a1*$^{Aga2/+}$ (Aga2) mouse, which recapitulates a moderate form of OI. Sc-RNAseq analysis of the cartilage growth plate revealed increased FGF signaling communication between perichondrial and growth plate tissues as well as increased expression of *Fgf21* and downstream targets (*Atf4* and *Ddit3*) in Aga2 perichondrial cells and pre-hypertrophic chondrocytes. FGF signaling and its connection to endoplasmic reticulum (ER) stress has not been explored in OI but has been extensively studied in achondroplasia and other growth plate disorders. Further, ER stress and the Unfolded Protein Response (UPR) in type I collagen expressing cells has been identified as a common underlying mechanism in several forms of OI and *Fgf21* is known to activate the UPR pathway thereby revealing a novel target for OI treatments. Pulmonary tissues in OI models have consistently displayed a histological emphysemous phenotype, however the origin of this and the effect on lung cell development and function remains unknown. Using sc-RNAseq data derived from young and adult Aga2 lungs, we found significantly decreased AT2 to AT1 cell transition (cells necessary for alveolar structure and gas exchange) in young Aga2 mice leading to an increased number of surfactant-producing AT2 cells in adults. Further, Aga2 lungs show increased myofibroblast numbers and matrix fibroblasts show increased EMT marker expression indicating a chronic damage response leading to an impaired response to tissue damage shown via ex-vivo lung bleomycin treatment. Clinical treatments specific to pulmonary complications in OI are non-existent and our results reveal the AT2 cell and increased surfactant production as novel targets to prevent the pulmonary insufficiency and early mortality observed in OI patients. Overall, our findings reveal novel cellular mechanisms behind extra-skeletal tissue abnormalities in OI and demonstrates how sc-RNAseq analysis of a rare disease model can uncover treatment targets that have not been identified using conventional methods.

S38. Disorders of bone pathology X-rayed using common and rare disease approaches

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 239. Low-dose infigratinib, an oral selective fibroblast growth factor receptor tyrosine kinase inhibitor, demonstrates activity in a preclinical model of hypochondroplasia

Authors:

C. Dambkowski[1], B. Demuynck[2], L. Loisay[2], L. Legeai-Mallet[2]; [1]QED Therapeutics, Inc., San Francisco, CA, [2]INSERM U1163, Imagine Inst., Paris Univ., Paris, France

Abstract:

Background: Fibroblast growth factor receptor 3 (*FGFR3*) gain-of-function mutations play a key role in achondroplasia (ACH) and hypochondroplasia (HCH). HCH is a less severe form of dwarfism than ACH, but is also caused by *FGFR3* gain-of-function mutations. HCH is characterized by disproportionate short stature and growth deficit affecting endochondral and intramembranous ossification. While various strategies are being tested, there are currently no approved therapies for HCH. We tested the hypothesis that the oral, selective FGFR tyrosine kinase inhibitor (TKI) infigratinib (BGJ398) could correct the HCH phenotype and improve endochondral and intramembranous ossification in a preclinical mouse model of HCH *Fgfr3*$^{N534K/+}$. **Methods:** The *Hch* mouse model expresses the most frequent human mutation *p.Asn540Lys* (*Fgfr3*$^{Asn534Lys/+}$) and exhibits mild dwarfism and most of the hallmarks of the human pathology. *Fgfr3*$^{N534K/+}$ mice received s.c. injections of infigratinib or vehicle control every 3 days (1 mg/kg) or daily (1 mg/kg) for 15 days (post-natal day [PND] 4-19) or 21 days (PND 3-24), respectively. **Results:** *Fgfr3*$^{N534K/+}$ mice treated with 1 mg/kg infigratinib every 3 days did not show obvious and significant modification of the dwarf phenotype. In contrast, *Fgfr3*$^{N534K/+}$ mice treated with 1 mg/kg infigratinib daily for 21 days showed a statistically significant increase in appendicular and axial skeletal measures. Length of the long bones was statistically significantly increased in *Fgfr3*$^{N534K/+}$ mice compared with *Fgfr3*$^{+/+}$ mice (tibia +3.18%, femur +3.16%, humerus +3.04%, ulna +2.94%, radius +3.01%). Treatment also modified the skull shape (skull width, skull height, nasal bone and naso-occipital length), the length of the mandible and skull base, as demonstrated by measurement of the foramen magnum (foramen magnum length +3.72%). Infigratinib treatment modified cartilage growth plate organization, in particular the hypertrophic chondrocyte area. Finally, high activation of the MAP kinase pathway due to the HCH missense *FGFR3* mutation was reduced by treatment, as revealed by immunolabelling of phosphorylated Erk1/2 proteins. **Conclusions:** Treatment with daily 1 mg/kg infigratinib improved the length and weight of *Fgfr3*$^{N534K/+}$ mice and significantly modified the axial and appendicular skeleton. We demonstrated in *Fgfr3*$^{N534K/+}$ mice that infigratinib is able to counteract the constitutive activation of FGFR3 due to the heterozygous *N540K* mutation localized in the tyrosine kinase domain of the protein. These results provide a rationale for targeting FGFR3 with a specific TKI for the treatment of children with HCH.

S39. Navigating the complex genetic landscape of neurodevelopmental disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 242. Functional genomics provide key insights to improve the diagnostic yield of hereditary ataxia

Authors:

Z. Chen[1], A. Tucci[2], V. Cipriani[2], E. Gustavsson[1], K. Ibanez[2], D. Zhang[1], L. Vestito[3], R. H. Reynolds[3], D. Smedley[2], H. Houlden[1], J. A. Botia[4], M. Ryten[3]; [1]Queen Square Inst. of Neurology, UCL, London, United Kingdom, [2]Queen Mary Univ. of London, London, United Kingdom, [3]Great Ormond Street Inst. of Child Hlth., UCL, London, United Kingdom, [4]Univ. of Murcia, Murcia, Spain

Abstract:

Hereditary ataxias (HA) are a group of clinically heterogeneous conditions characterised by progressive incoordination from cerebellar dysfunction. Despite the critical mass of 300 associated genes being discovered and the improvements in functional genomic annotation, up to 80% of patients remain molecularly undiagnosed even following high-depth whole genome sequencing (WGS). Our study aimed to leverage multi-omics data to further characterise the genetic architecture of HA and to improve their diagnostic yield.

We generated 294 genic features, capturing information about a gene's structure; genetic variation; tissue-specific, cell-type-specific and temporally-relevant expression and protein products. Using modified National Genomic Test Directory age-of-onset information (currently in clinical practice in NHS England), we categorised 318 HA-linked genes as childhood-onset, adult-onset and those overlapping both. We then compared these individual genomic features across gene categories and collectively through unsupervised learning. We verified our findings using patient data from the 100,000 Genomes Project.

By comparing these properties, we demonstrated: (i) an unexpectedly high short tandem repeat (STR) density within childhood-onset genes suggesting that we may be missing pathogenic repeat expansions in this cohort; (ii) cell-type-specific gene expression differentiates childhood- and adult-onset ataxias with CNS glial-specific expression in childhood-onset and Purkinje cell-specific expression in overlap-onset genes; (iii) significant similarities in annotation across the groups using unsupervised analysis suggesting adult- and childhood-onset patients should be screened using a common gene set.

To test these findings, we found that both childhood- and adult-onset ataxia patients (n=553 and 1760 respectively) from the 100,000 Genomes Project exhibited a trend for higher repeat sizes genotyped by ExpansionHunter using WGS data even at naturally-occurring STRs in known ataxia genes compared with neurologically-normal participants (n=6078). This implies a role for STRs in the disease mechanism of HA and prioritises candidates to explore in an unsolved cohort. We also assessed the burden of potentially pathogenic variants among childhood-onset genes in these adult-onset HA patients and vice versa. This demonstrated a significantly higher burden of rare loss-of-function variants in the childhood-onset HA genes among adult-onset HA patients.

Our analysis suggests a modified testing strategy for HA and highlights genic features of interest to explore in an unsolved cohort to increase the diagnostic yield.

S39. Navigating the complex genetic landscape of neurodevelopmental disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 243. Contribution of copy number variants to schizophrenia in East Asian populations

Authors:

Q. Feng[1], B. Thiruvahindrapuram[2], D. Howrigan[3], Stanley Global Asia Initiatives, S. Qin[4], J. Sebat[5], T. Ge[1], H. Huang[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Peter Gilgan Ctr. for Res. and Learning, Toronto, ON, Canada, [3]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, Cambridge, MA, [4]Shanghai Jiao Tong Univ., Shanghai, China, [5]Univ. of California, San Diego, La Jolla, CA

Abstract:

Studies have established an important role for copy number variants (CNVs) in the etiology of schizophrenia (SCZ) by showing higher effect sizes of rare CNVs (OR of 4-100) than common variants (OR<1.3). However, to date, most studies were conducted in individuals of European descent (EUR), with only a handful of small-scale studies in other populations. Thus the extent to which CNVs contribute to schizophrenia in non-EUR populations is unclear. Here we report findings from the largest CNV study in individuals of East Asian ancestry (EAS), including 17,749 SCZ cases and 19,252 controls. After extensive quality control harmonizing datasets across various arrays, we observed a significant enrichment of genome-wide rare CNV burden in SCZ cases (OR=1.023 [1.013-1.032], P=1E-6) driven by gene-overlapping CNVs with length < 100kb and frequency < 1%. The CNV burden remains significant even after we excluded all previously reported SCZ-associated CNVs (OR=1.018 [1.008-1.028], P=1E-4), suggesting there exists SCZ-associated CNVs in EAS that were not discovered in previous EUR studies of larger sample sizes. Interestingly, the CNV burden in SCZ cases of EAS is slightly smaller than but statistically comparable with that in EUR (OR=1.028 [1.022-1.034] from PGC), likely because most arrays were optimized for EUR and thus had better CNV calling accuracy in EUR. Breakpoint-level CNV association analysis in EAS replicated previously reported CNV loci, including 22q11.21 del, NRXN1 del, and 16p11.2 dup at nominal significance. Interestingly these top CNV associations in EUR had larger effect sizes than those in EAS although the differences were not of statistical significance. For example, the 22q11.2 deletion had OR=3.7 [1.42-10.34] in EAS, compared with OR=67.7 [9.3-492.8] in EUR from PGC. We found 3 novel CNV-SCZ associations using this EAS samples and through a meta-analysis with EUR samples (16p13.11, 17q22.1 and 1p21.3), driving the total number of SCZ-associated CNVs from 3 (from PGC) to 6. Among them, 16p13.11 (OR= 5.34 [2.22-12.83]) was found associated with intellectual disability and autism in previous studies. In summary, we for the first time studied the CNVs contribution to SCZ across ancestries with large-scale EAS sample. Comparing CNVs discovered in EAS and EUR, we found EAS hosts a smaller CNV burden in SCZ cases and has smaller effects for many known CNV associations, although none of the differences reached statistical significance. Despite a smaller sample size, we identified novel SCZ-associated in EAS not discovered in EUR, suggesting the importance of future large-scale studies to study the CNV contribution to psychiatric disorders across ancestries.

S39. Navigating the complex genetic landscape of neurodevelopmental disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 244. *In vivo* functional screening in the mouse brain uncovers autism-relevant enhancers

Authors:

J. Lambert[1], K. Cichewicz[1], L. Su-Feher[1], T. Warren[1], J. Haigh[2], S. Morse[1], A. Santos[1], E. Castillo Palacios[1], J-M. Hanners[1], A. Nord[1]; [1]Univ. of California, Davis, Davis, CA, [2]Leeds Becket Univ., Leeds, United Kingdom

Abstract:

Determining sequence variants that contribute to risk of autism spectrum disorder (ASD) remains a primary goal towards understanding the etiology of this neurodevelopmental condition. Whole genome sequencing has uncovered *de novo* and rare variants that may be associated with ASD, the majority of which lie in noncoding DNA. Predicting which of these non-coding variants, if any, contribute to ASD risk is a major challenge. Among noncoding DNA sequences, *cis* regulatory enhancers have emerged as critical drivers of cell-specific gene expression and of complex genetic disease risk. Studies of rare and *de novo* enhancer mutations in ASD have identified risk burden and individual examples have been shown to impact regulatory function. While *in silico* risk prediction continues to improve, there is a major need for experimental screening to identify which enhancers and which variants may matter. Towards this goal, we used the self-transcribing active regulatory region sequencing (STARR-seq) massively parallel reporter assay (MPRA) strategy, testing a library of candidate enhancers found to be mutated in individuals with ASD or their unaffected siblings. We cloned candidate enhancers including 900 bp flanking 1,015 *de novo* variants and their reference alleles into the 3' UTR of an EGFP reporter gene under the control of a minimal promoter. We then generated a library of rAAV and transduced neonatal mouse brain. Among enhancer sequences interrogated, we found a 1.31-fold enrichment of active enhancers from candidate regions from autism probands compared to unaffected siblings. We identified altered enhancer activity levels between a number of variant and reference alleles that we are experimentally validating, including mutations located within introns of *DOCK1* and *SEPT9* . Our work lays the foundation for using genomics-based assays for testing enhancer function to determine potential ASD causal variants based on *in vivo* functional relevance and is generally applicable across complex genetic diseases.

S39. Navigating the complex genetic landscape of neurodevelopmental disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 245. Single-cell transcriptomic benchmarks of human iPSC-derived neuronal cultures and their implications for modeling neurological disorders

Authors:

D. Gao[1,2,3], D. J. C. Tai[1,2,3,4], C. de Esch[1,2,3,4], V. A. Christine[2,5], M. Kiana[1,2,3,4], X. Nuttle[1,2,3], B. Currall[1], K. O'Keefe[1], N. Burt[1], R. Bhavsar[1,2,3], S. Erdin[1,2,3], J. F. Gusella[1,2,3], R. Yadav[1,2,3,4,5], D. Bragg[2,5], M. Talkowski[1,2,3,4,5]; [1]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [2]Dept. of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [5]The Collaborative Ctr. for X-Linked Dystonia-Parkinsonism, Massachusetts Gen. Hosp., Charlestown, MA

Abstract:

Large single-cell RNA (scRNA) studies contain intrinsic variances (IVs) from cell-type differences and extrinsic variances (EVs) introduced by technical confounding factors. In iPSC-derived neuronal modeling, these IVs and EVs can result from differentiation protocols, batch variation, and differences between laboratories, but little is known about how they influence scRNA interpretation. Here, we generated 60 scRNA libraries for iPSC-derived neuronal lineages, including neural stem cells (NSCs), NGN2-induced neurons (iNs), microglia, and cerebral organoids (COs), made across multiple laboratories using different background lines and differentiation protocols in multiple batches. Unsupervised clustering successfully grouped each lineage, irrespective of EVs. However, EVs significantly confounded the cellular heterogeneity within each lineage. Among NSCs, iNs and COs, iNs demonstrated 13% total variance contributed by all EVs combined, compared to 18% and 38% in NSCs and COs, respectively. After correcting for the EVs, iNs were more homogenous than NSCs, with 27.4% less remaining variance accounted for by IVs. These results suggest iNs have the highest reproducibility and homogeneity of these lineages. When we compared the 200 most significant lineage signature genes to the annotated cell types in human postmortem brains from UCSC Cell and Allen Brain, 50% of NSCs matched the annotated progenitors at various fetal development stages (FDR < 0.001, hypergeometric test), while 78% of cells in COs matched excitatory neuron (ExN), inhibitory neuron (InhN), and astrocyte signatures in the brains. In contrast, iNs did not resemble any annotated neuron type based on the top 200 signatures, although expressing ExN markers *SNAP25* and *SYT4* sporadically. With EV-derived effects and IV-derived cell-type diversity established, we examined their impacts on neuronal disorder mechanisms. Using a study of the ~565kb 16p11.2 copy number variant effect in autism as an example, eight organoids were created with 16p11.2 deletion, duplication, and control lines in the same batch and with identical methods. When we investigated the unique co-expression networks defining each cell type, co-expression modules positively correlated with 16p11.2 gene dosage were observed in all cell types. However, a subtype of InhNs marked by *CALB2*+ co-expression illuminated a module negatively correlated with 16p11.2 dosage (Pearson's r = -0.97, p < 2.2e-16) and was enriched for neuron projection genes. These analyses emphasize the critical factors confounding scRNA interpretation and illuminate the value of cell-type-specific studies from mixed organoid cultures.

S39. Navigating the complex genetic landscape of neurodevelopmental disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 246. Dosage imbalance of the chromatin remodeler *CHD1L* contributes to the 1q21.1 CNV-associated mirrored neurodevelopmental phenotypes

Authors:

M. Lemée, M-N. Loviglio, C. Keime, T. Ye, P. Tilly, C. Weber, O. Wendling, H. Jacobs, J. D. Godin, C. Golzio; IGBMC, Illkirch-Graffenstaden, France

Abstract:

Recurrent reciprocal distal 1q21.1 Deletions and Duplications have been found in individuals with syndromic autism. Variable phenotypes have been reported, including congenital heart defects, autism, schizophrenia, head circumference and height defects. The deletion is associated with microcephaly and short stature whereas the reciprocal duplication is associated with macrocephaly and carriers tend to be in the upper height percentiles. We modeled the 1q21.1 Dup by overexpressing each of the eight human genes within the 1q21.1 CNVs in zebrafish. Strikingly, we found that the overexpression of *CHD1L* induces increased number of post-mitotic neurons at 2 days post-fertilization (dpf), resulting in macrocephaly at 5dpf (20% increase of brain volume). Consistently, suppression of the zebrafish ortholog of *CHD1L* leads to microcephaly (25% decrease of brain volume) and apoptosis at 5dpf. We also observed a mirrored effect on larval body length, a readout for height. Transitioning zebrafish data into mammalian systems yielded consistent phenotypes. We observed mirrored effects on the number of mature Tbr1+ neurons one day after *in utero* electroporation of E13 mouse embryos with either *Chd1l* construct to mimic an overexpression or short hairpin RNAs targeting *Chd1l* to mimic an acute knockdown. We further explored the role of *CHD1L* during early neurogenesis by generating isogenic, control and *CHD1L* loss-of-function, human induced pluripotent stem cells. The lines were differentiated into neuronal progenitor cells (NPC) and subjected to RNA-seq and ATAC-seq. These approaches revealed that loss of *CHD1L* affects chromatin accessibility and expression levels of genes involved in neuronal differentiation and synaptogenesis including susceptibility autism genes such as *UNC5D* and *DPP6*. We also found that absence of *CHD1L* leads to microcephalic-like organoids and to a significant decreased of TUJ1-positive neurons layer in cortical organoids. Last, we asked whether loss of *CHD1L* might be sufficient to cause phenotypes in humans. A role for *CHD1L* dosage changes is consistent with autism in both a shorter atypical 323 kb deletion at 1q21.1 encompassing *CHD1L* only and a homozygous missense variant of unknown significance in *CHD1L* (p.Arg392His) for which we confirmed its pathogenicity by zebrafish complementation assay. Taken together, our data obtained in zebrafish, mice, human NPCs and organoids indicate that the loss of *CHD1L* perturbs early stages of neuronal differentiation and suggest that *CHD1L* dosage changes contribute to the 1q21.1 CNV mirrored neuroanatomical phenotypes.

S39. Navigating the complex genetic landscape of neurodevelopmental disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 247. Bi-allelic pathogenic variants in *TMEM147* cause moderate to profound intellectual disability with dysmorphic facial features

Authors:

Q. Thomas[1], M. Motta[2], T. Gauthier[3], M. Zaki[4], A. Ciolfi[5], J. Paccaud[6], F. Girodon[7], O. Boespflug-Tanguy[8], T. Besnard[9], J. Kerkhof[10], H. McConkey[11], A-S. Denomme-Pichon[12], B. Cogné[13], L. Rodan[14], R. Abou Jamra[15], A. Jackson[16], S. Banka[17], M. Mc Entagart[18], A. Saggar[19], J. Gleeson[20], D. Sievert[21], H. Bae[22], K. Kwon[23], G. Seo[24], B. Hee Lee[25], H. Lee[26], A. Saeed[27], N. Anjum[27], H. Sheema[27], S. Alawbathani[28], I. Khan[29], J. Pinto Basto[30], J. Teoh[31], J. Wong[31], U. Bin Mohamad Sahari[31], H. Houlden[32], K. Zhelcheska[32], J. Delanne[33], M. Pannetier[34], C. Philippe[6], L. Faivre[33], S. Odent[35], A. Bertoli Avella[28], C. Thauvin-Robinet[36], B. Sadikovic[37], B. Reversade[38], R. Maroofian[39], J. Govin[3], M. Tartaglia[40], A. Vitobello[6]; [1]Genetics Ctr., FHU-TRANSLAD and GIMI Inst., Dijon Univ. Hosp., Dijon, France, [2]Genetics and Rare Diseases Res. Div., Ospedale Pediatrico Bambino Gesù, Rome, Italy, [3]Inst. for Advanced Biology, Ctr. de Recherche UGA, INSERM U1209, CNRS UMR 5309, Grenoble, France, [4]Natl. Res. Ctr., Cairo, Egypt, [5]Univ. of Torino, Torino, Italy, [6]Inserm UMR1231 team GAD, Univ. of Burgundy and Franche-Comté, Dijon, France, [7]Hematology Deparment, Dijon Univ. Hosp., Dijon, France, [8]Hosp. Robert-Debré, APHP, Paris, France, [9]CHU de Nantes Hôtel-Dieu, NANTES CEDEX 1, France, [10]London Hlth.Sci. Ctr., London, ON, Canada, [11]Verspeeten Clinical Genome Ctr., London Hlth.Sci. Ctr., London, ON, Canada, [12]Inserm UMR1231 team GAD, University of Burgundy and Franche-Comté, France, [13]Université de Nantes, CNRS, INSERM, l'institut du thorax, Dijon, France, [14]Genetics and Metabolism, Div. of Genetics and Genomics, Boston Children's Hosp., Boston, MA, [15]Leipzig Med. Univ., Leipzig, Germany, [16]Div. of Evolution & Genomic Sci., Sch. of Biological Sci., Faculty of Biology, Med. and Hlth., Univ. of Manchester, Manchester, United Kingdom, [17]Univ. of Manchester, Sale, United Kingdom, [18]St George's Hosp., London, United Kingdom, [19]Intl. Gene Clinic, London, United Kingdom, [20]Univ California, La Jolla, CA, [21]Univ California, [22]Med. Genetics Ctr., Asan Med. Ctr. Children's Hosp., Seoul, Korea, Republic of, [23]3billion, Seoul, Korea, Republic of, [24]3billion, Inc, Seoul, Korea, Republic of, [25]Dept. of Pediatrics, Asan Med. Ctr. Children's Hosp., Univ. of Ulsan Coll. of Med.,, Seoul, Korea, Republic of, [26]Dept. of Pediatrics, Asan Med. Ctr. Children's Hosp., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [27]Children's Hosp. of Lahore, Lahora, Pakistan, [28]CENTOGENE GmbH, Rostock, Germany, [29]CENTOGENE GmbH, Singapore, Singapore, [30]Centogene, Rostock, Germany, [31]Genome Inst. of Singapore, A*STAR, Singapore, Singapore, [32]Dept. of Neuromuscular Disease, UCL Queen Square Inst. of Neurology and The Natl. Hosp. for Neurology and Neurosurgery, London, United Kingdom, [33]Genetics Ctr., FHU-TRANSLAD, Dijon Bourgogne Univ. Hosp., Dijon, France, [34]Service de Génétique Médicale, CHU Rennes, Rennes, France, [35]CHU de Rennes Hosp SUD, Rennes CDX 2, France, [36]Genetics Ctr., FHU-TRANSLAD, Dijon Bourgogne Univ. Hosp., Dijon, Dijon, France, [37]LHSC, London, ON, Canada, [38]Genome Inst. of Singapore, Singapore, Singapore, [39]Dept. of Neuromuscular disorders, UCL Inst. of Neurology, London, United Kingdom, [40]Ospedale Pediatrico Bambino Gesù, Rome, Italy

Abstract:

The transmembrane protein TMEM147 has been recently shown to localize at nuclear envelope, where it anchors lamin B receptor (LBR) to the inner membrane, and at the endoplasmic reticulum (ER) acting as an important component of the ribosome-associated translocon complex, facilitating the insertion and translocation of newly synthesized proteins at the ER membrane. Over the years, articles related to *TMEM147*'s partners have shown their paramount importance in early embryonic development, mesoendoderm induction, right-left patterning, neural patterning and final patterning of the vertebrate central nervous system making *TMEM147* a strong candidate for neurodevelopmental disorders. However, the role of TMEM147 in human disease has not been characterized so far. Through international data sharing, we identified 22 individuals from 14 unrelated families with bi-allelic *TMEM147* pathogenic variants, including missense, splice site, and nonsense of frameshift variants, who displayed congruent clinical features including coarse facies, intellectual disability, developmental delay and behavioral problems. *In silico*, missense variants were predicted to cause steric clashes with TMEM147 itself, Nicalin or TMCO1, additional translocon accessory factors. *In vitro*, engineered missense mutants induced protein instability via autophagy/lysosomal-mediated degradation, without altering ER localization. Furthermore, TMEM147-deficient cells showed CLIMP-63/CKAP4 and RTN4/NOGO upregulation with a concomitant reorientation of the ER in patient-derived fibroblasts. Abnormal nuclear segmentation and chromatin compaction were observed in approximately 20% of neutrophils. Finally, co-expression analysis revealed significant correlation with neurodevelopmental (NDD) genes in the brain. In conclusion, we provide genetic and functional evidence that biallelic loss-of-function variants in *TMEM147* are associated with syndromic intellectual disability due to ER-translocon and nuclear organization dysfunction in humans.

The full list of the authors include also "Aymeric Masson[1],Eva Trochu[9], Virginie Vignard[13], Fatima El It[1], Yannis Duffourd[1], and Ange-Line Bruel[1]

S40. Polygenic prediction of complex disease

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 250. Genome-wide meta-analysis identifies novel loci and improves disease prediction of Age-related Macular Degeneration

Authors:

W. He, X. Han, J. Ong, P. Gharahkhani, S. Macgregor; QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Abstract:

Age-related macular degeneration (AMD), a progressive degeneration of the macular, is the leading cause of central vision loss in aged people. Research showed there were approximately 196 million AMD cases in 2020, and it is expected to reach 288 million in 2040. While previous efforts uncovered several genes contributing to advanced AMD, identifying more genetic variants will help to improve the prediction of disease risk and may provide new insights into disease aetiology. Here we conducted a meta-analysis of several GWAS: International AMD Genomics Consortium AMD-2016 GWAS (16,144 advanced AMD cases and 17,832 controls), AMD-2013 GWAS (17,181 cases and 60,074 controls), UK Biobank AMD GWAS (8864 cases and 140,685 controls), Genetic Epidemiology Research on Aging study (4017 AMD cases and 14,984 controls), Finngen R6 AMD GWAS (4645 cases and 243,951 controls) and an early-AMD study (11 sources including 14,034 cases and 91,214 controls). We employed a multi-trait approach which dealt with sample overlap. We identified 13 novel genome-wide significant independent AMD risk loci near or within several genes which have previously been linked to glaucoma risk (HIC1, ME3). We constructed a polygenic risk score (PRS) for AMD based on these novel loci and previously suggested AMD loci, and examined the genetic risk of people in non-overlapping data from the Canadian Longitudinal on Aging Study (733 AMD cases and 20,487 controls). The PRS showed significantly improved prediction accuracy for AMD ($p = 0.005$). We also established a new disease prediction model for AMD patients based on PRS, age, sex, smoking status and top 10 genetic principal components (AUC = 0.79). We assessed the PRS in AMD cases from the Canadian Longitudinal on Aging Study and found PRS was associated with the age of disease onset. We found the cumulative incidence of AMD by age 80 in the high (10%) PRS group was more than four times higher (19% ± 2% vs 5% ± 2%) than in the low (10%) PRS group. Our findings improve the knowledge of the genetic architecture of AMD and help achieve better accuracy in AMD prediction.

S40. Polygenic prediction of complex disease

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 251. Incorporating genetic information into clinical prediction models: Assessment of technical considerations and value in Crohn's disease

Authors:

T. Morley[1], D. Wilimitis[1], H. Lee[2], K. Choi[2], M. Ripperger[1], C. Walsh[1], J. Smoller[3], J. Kang[1], D. Ruderfer[1,4]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Mass Gen. Brigham, Boston, MA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Vanderbilt Univ., Nashville, TN

Abstract:

Introduction: Genome-wide data continues to be more accessible and more frequently considered in the clinic, with an increasing number of studies returning such data to patients. However, there remain questions around its value for clinical risk prediction, and how best to integrate it with existing clinical risk models. We assess the value of genetic data in disease prediction via various methods of combining a polygenic risk score (PRS) with longitudinal health care data. We examine Crohn's disease due to quality of phenotyping algorithm, prevalence and well described genetic architecture.
Methods: Models were developed at Vanderbilt University Medical Center (VUMC) and Mass General Brigham (MGB). Crohn's Disease was defined based on a PheKB algorithm using diagnostic codes and medications. We compared performance of incorporating PRS as an additional feature (early fusion) or training a model combining PRS and the output of a clinical model (late fusion). We used nested cross validation to view performance across our entire sample. Our PRS was calculated with PRS-CS, on the most recent IBD GWAS. Performance was measured using area under the receiver operator curve (AUROC) and area under the precision recall curve (AUPRC).
Results: Across sites 4,786 patients met criteria for Crohn's disease. Among those with genetic data, there were 243 cases and 77,105 controls. PRS alone showed AUROC of 0.68 at both sites and AUPRC of 0.011 (VUMC) and 0.008 (MGB) compared to a model based on demographics (age, sex, race) which had AUROC of 0.74 (VUMC) and 0.69 (MGB) and AUPRC of 0.01 (VUMC) and 0.007 (MGB). Fusing PRS with demographics increased AUROC and AUPRC at VUMC but did not affect performance at MGB. Clinical models had similar AUROC trained in or out of our biobank, with higher AUPRC in the biobank. As we increased model complexity, we saw substantial performance increases. Adding PRS via late fusion had similar AUROC and AUPRC to the clinical model. We saw increased AUPRC using early fusion, going from 0.214 to 0.233 at VUMC and 0.51 to 0.541 at MGB.
Discussion: This work explored methodological considerations when integrating genetic and clinical data for risk prediction. We show that the largest performance gains were from model complexity. When incorporating PRS into a clinical risk prediction model, we find that including it as an additional feature as opposed to combining with the clinical probability provides opportunity for better performance, suggesting that PRS may improve performance through nonlinear interactions with clinical features. These results have implications in determining best practices for future efforts to incorporate genetic data in risk prediction.

S40. Polygenic prediction of complex disease

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 252. Impacts of individual uncertainty in polygenic risk score estimation on lung cancer risk stratification and prediction

Authors:

X. Wang[1], Z. Zhang[1], Y. Ding[2], L. Su[1], L. Mucci[3], C. Amos[4], D. Christiani[5]; [1]Harvard Univ., Boston, MA, [2]Univ. of California, Los Angeles, Los Angeles, CA, [3]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Baylor Coll. of Med., Houston, TX, [5]Harvard Med. Sch., Boston, MA

Abstract:

Although polygenic risk scores (PRSs), which summarize genetic risk at the individual level, have been widely studied in many diseases and traits, individual-level uncertainty in PRSs remains largely unexplored. Here, we estimated the uncertainty of lung cancer PRS at the individual level and evaluated its impact on subsequent risk stratification and prediction. Lung cancer PRS with individual credible sets were constructed via two approaches - a GWAS-based bootstrapping method and a Bayesian method among 18,146 lung cancer cases and 12,894 cancer-free controls of European ancestry in the International Lung Cancer Consortium (ILCCO). Considerable variance in PRS point estimates at the individual level was observed for both methods, with an average standard deviation (s.d.) of 0.12 (95% CI = 0.09-0.15) for the GWAS-based bootstrapping method and s.d. = 0.88 (95% CI = 0.68-1.11) for the Bayesian method. The bootstrapping method classified only 2.4% and 1.5% of the target population into the low and high-risk strata with certainty, while the Bayesian approach was unable to find any of the individuals with certainty at the population threshold t=10th and 90th percentiles with a credible level p=95%. Taking individual uncertainty into account, a stronger relative risk of lung cancer was observed for individuals in the top decile of the PRS under the confident stratification rule (OR=2.92,95% CI=2.26-3.78, p-value=$2.5*10^{-16}$) compared to the naïve stratification method based on PRS estimate alone (OR=2.20, 95% CI=1.99-2.45, p-value=$1.6*10^{-50}$). In stratified analysis, we observed similar changes in effect sizes across all subgroups, with larger discrepancies in small cell lung cancer patients and never smokers. Our findings characterize the individual level uncertainty in lung cancer PRS in European population and this uncertainty can potentially impact lung cancer risk stratification and prediction. It is worthy of serious consideration and discovery in individual-level PRS uncertainty before it being equitably and reliably utilized in clinical setting.

S40. Polygenic prediction of complex disease

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 253. Evaluation of type 2 diabetes polygenic risk scores in prospective longitudinal cohort study in Korean population

Authors:

N. Kim[1], H. Lee[2], S. Kim[3], Y-J. Kim[3], H. Lee[4], S. Kwak[4], S. Lee[1]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Soonchunhyang Univ. Seoul Hosp., Seoul, Korea, Republic of, [3]Dept. of Clinical Epidemiology and Biostatistics, Asan Med. Ctr., Seoul, Korea, Republic of, [4]Dept. of Internal Med., Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Abstract:

Polygenic risk score (PRS) is an important tool to predict individual genetic susceptibility to type 2 diabetes mellitus (T2DM). However, current PRS research is largely limited to a disease prediction on cross-sectional data, and the potential clinical utility of T2DM PRS in predicting incident T2DM is not fully evaluated. In this study, we first construct East Asian T2DM PRS using large biobank data of Korea and Japan. We carried out genome-wide association analysis and meta-analysis to construct GWAS summary for PRS training with the total sample size being 269,487, and applied penalized regression framework-Lassosum and Bayesian regression framework-PRScs to construct PRS. Second, using a prospective, community-based Asan/Ansung cohort of the Korean Genome and Epidemiology Study (KoGES) with 16 years of follow-ups, we demonstrate additional value in PRS for predicting the incidence of T2DM and the progression from pre-diabetes to T2DM and to the insulin prescription. Specifically, prediction performance of PRS was evaluated in a series of models, including family history, physical measurements, and clinical factors with and without PRS. Our results showed that T2DM PRS could predict not only the progress from non-diabetes to T2DM, but also non-diabetes to prediabetes and prediabetes to T2DM. Hazard ratios that compared the top decile of PRS group with the middle PRS group (10th to 90th percentiles) were 1.60 (95% CI: 1.07-2.40) for the overall risk of T2DM in non-diabetes patients, 1.59 (95% CI: 1.31-1.94) for the progression to pre-diabetes from non-diabetes, and 1.76 (95% CI: 1.46-2.14) for the progression to T2DM from pre-diabetes. We also found that T2DM patients in the higher percentile PRS group were more likely to be prescribed insulin. Each standard deviation increased in the standardized PRS was associated with 90% higher odds of being prescribed insulin (odd ratio = 1.90, 95% CI:1.56-2.31), p-value= 1.47e-10). The comparison of the prediction models of the incidence of T2DM with and without PRS showed that PRS improved the overall prediction accuracy of AUC by 4.51-11.44% across the considered sequential models. By analyzing prospective cohort study data, we demonstrated that T2DM PRS could help identifying high-risk groups to progress prediabetes and T2DM, and individuals with severe T2DM who need to be prescribed insulin.

S40. Polygenic prediction of complex disease

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 254. Integrating large scale genetic and clinical information to predict cases of heart failure

Authors:

K-H. Wu[1], W. Hornsby[2], X. Yu[1], S. Graham[1], M. Mathis[3], N. Douville[3], D. Amare[4], Global Biobank Meta-analysis Initiative (GBMI), I. Surakka[1], X. Shi[1], C. Willer[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Michigan Med., Ann Arbor, MI, [4]Univ. of Adelaide, Adelaide, Australia

Abstract:

[Background] Heart disease is the leading cause of death globally and earlier initiation of treatment could mitigate disease progression. Multiple efforts have been made using genome-wide association studies (GWAS) or electronic health records (EHR) to identify individuals at high risk of heart failure (HF). However, integrating both sources using novel natural language processing (NLP) technique and large scale global genetic study into heart failure prediction models has not been evaluated. **[Objectives]** The study aimed to improve the accuracy of HF prediction by integrating GWAS- and EHR-derived risk scores. **[Methods]** We performed the largest HF GWAS to date within the Global Biobank Meta-analysis Initiative, which includes 1.3 million samples (67,049 cases; 5%) from 13 biobanks across the world, to create a polygenic risk score (PRS). To extract information from high-dimensional EHR, we treated diagnosis codes as 'words' and applied NLP on the data. NLP was used to learn code co-occurrence patterns and extract 350 latent phenotypes (low-dimensional features) to represent (autofill) the rest of the 29,346 EHR codes. Next, we regressed HF on the latent phenotypes in an independent cohort and the coefficients were used as the weights to calculate clinical risk score (ClinRS). Model performances were compared between baseline (age and sex) model and three models with risk scores added: 1) PRS, 2) ClinRS, and 3) PRS+ClinRS, using 10-fold cross validated Area Under the Receiver Operating Characteristic Curve (AUC). **[Results]** Our results show that PRS and ClinRS separately are able to predict HF outcomes significantly better than the baseline model, up to 7 years prior to HF diagnosis. Higher AUC (95% CI) were observed in the PRS model (0.77 [0.75-0.80]) and ClinRS model (0.79 [0.76-0.81]), compared to the baseline model (0.71 [0.69-0.74]). Moreover, by including both PRS and ClinRS in the model, we achieved superior performance in predicting HF up to 10 years prior to HF diagnosis (AUC: 0.79 [0.77-0.82]), three years earlier than using single risk predictor alone. **[Conclusions]** We demonstrate the additive power of integrating GWAS- and EHR-derived risk scores to predict HF cases prior to diagnosis. Clinical application of this approach may allow identification of patients with higher susceptibility to HF and enable preventive therapies to be initiated at an earlier stage.

S40. Polygenic prediction of complex disease

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 255. Investigating genetic drivers of cardiovascular disease variability in admixed individuals

Authors:

G. Tietz, N. N. Shah, T. Tan, E. G. Atkinson; Baylor Coll. of Med., Houston, TX

Abstract:

Cardiovascular disease (CVD) is the leading cause of death in the United States, and in 2021 accounted for 20% of all deaths. To mitigate CVD and inform patient care, genetic testing is becoming increasingly common to identify those at risk and enable them to take preventive measures. However, current prediction models are missing key information as they evaluate variant pathogenicity independently and do not account for the interaction between variants, otherwise known as epistasis. Further, these models do not consider the ancestral background of the variant, even though genetic variants are known to exhibit different effects across populations. Given the complexity of epistasis and ancestral factors, there remain substantial challenges to understanding the genetic drivers of CVD across diverse populations.

We use differential penetrance among ancestries to investigate the role of epistasis and ancestral genomic background on a CVD pathogenic variant. Murdock *et al* discovered that the *LPA* variant rs3798220 known to increase CVD in European ancestry individuals was not strongly associated with CVD for those in the Hispanic community. This variant is included in a HeartCare CVD panel and is considered clinically actionable, but does not appear to be predictive for all populations. To increase the transferability of genetic testing and allow for precision-medicine approaches beyond European cohorts, it is critical to determine the genetic drivers that explain the variable penetrance of the rs3798220 variant.

Here, we identified whether ancestral haplotypes surrounding rs3798220, or an additional variant enriched in the Hispanic population explained the decreased CVD risk. We leveraged data from the All of Us Research program, housing the largest number of Hispanic individuals in a biobank to-date, to assess the influence of local ancestry patterns surrounding rs3798220. We additionally conducted pairwise epistatic testing of rs3798220 across the genome and employed the *Tractor* pipeline to evaluate if this variant was modified by other loci.

Together, we demonstrate the impact of local ancestry and epistasis on the clinically actionable variant rs3798220. On a foundational level, this contributes to our understanding of epistasis across diverse human populations and why individuals with the same genotype may display differing phenotypes. In a clinical context, this refines CVD risk prediction for rs3798220 in historically underserved communities for whom genetic tests currently underperform. Our novel ancestry-informed approach can be applied broadly, giving it potential to improve precision-medicine across populations and complex traits.

S41. Populations evolving: Modeling genetic variation to understand evolutionary processes

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 258. Inference of coalescence times and allele ages using deep neural networks identifies signatures of natural selection

Authors:

Z. Tsangalidou, J. Nait Saada, M. Stricker, P. Palamara; Univ. of Oxford, Oxford, United Kingdom

Abstract:

Accurate inference of time to the most recent common ancestor (TMRCA) between pairs of individuals and of allele ages is key in several population genetic analyses. We developed CoalNN, a likelihood-free approach which uses deep convolutional neural networks to predict pairwise TMRCAs and allele ages from sequencing, SNP array, or imputed genotype data. CoalNN is trained through simulation and can be adapted through transfer learning to different evolutionary parameters, such as demographic history.
We compared CoalNN to other available methods in several simulated scenarios and observed consistently improved performance for TMRCA and allele age prediction. We used saliency maps and input perturbation analyses to demonstrate that CoalNN implicitly relies on combinations of features, such as minor allele frequency (MAF) and haplotype sharing, that require complex probabilistic modeling in other approaches.
We then applied CoalNN to 2,504 samples from 26 populations in the 1,000 Genome Project to infer the ages of ~80 million variants. Our estimates were highly correlated with other recent estimates (e.g. R=0.3, SE=0.07 with GEVA; R=0.67, SE=0.11 with Relate). We analyzed the distribution of inferred frequency-stratified allele ages, finding substantial inter-population variation, reflecting heterogeneous demographic histories. We observed that deleterious alleles, annotated using Polyphen2 and SIFT, are on average younger than neutral alleles of the same frequency, consistent with the action of negative selection.
Finally, we MAF-stratified and quantile normalized CoalNN's predicted allele ages to construct genome-wide annotations capturing the signature of past negative selection. We used LD-score regression and summary association statistics from 63 independent complex traits and diseases (average study N=314k) to estimate the effect of the CoalNN negative selection annotation on trait heritability. Conditioned on several other functional and evolutionary annotations from the Baseline-LD model (Gazal et al. 2017), the CoalNN annotation yielded larger effect estimates than a previous annotation based on allele ages (ARGWeaver τ^*=-0.147; SE=0.005; CoalNN τ^*=-0.218, SE=0.018). We further stratified this signal by ancestry, observing that although association statistics for these traits were computed in predominantly European samples, allele age estimates from non-European groups provided the largest contribution to the estimated effects on trait heritability.
Overall, these results underscore the efficacy of using deep neural networks in analyses involving genealogical relationships in large genomic data sets.

S41. Populations evolving: Modeling genetic variation to understand evolutionary processes

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 259. Genetic evidence for ancient population shifts and migrations in Central and Southern California

Authors:

N. Nakatsuka[1,2], B. Holguin[3], J. Sedig[4], P. E. Langenwalter[5], D. J. Kennett[3], R. Pinhasi[6], J. Punzo Diaz[7], J. R. Johnson[8], D. Reich[4]; [1]New York Genome Ctr., New York, NY, [2]Harvard Med. Sch., Boston, MA, [3]Univ. of California, Santa Barbara, Santa Barbara, CA, [4]Harvard Univ., Cambridge, MA, [5]Biola Univ., La Mirada, CA, [6]Univ. of Vienna, Vienna, Austria, [7]Inst. Natl. de Antrpología e Historia, Morelia, Mexico, [8]Santa Barbara Museum of Natural History, Santa Barbara, CA

Abstract:

California, prior to European contact, harbored more linguistic diversity than all of Europe, and studies of language relationships, archaeological evidence, and traditional oral histories have led to many hypotheses about movements of people over the minimum of 13,000 years this region has been occupied. In the absence of ancient DNA, it has not been possible to test directly alternative hypotheses about the spread of groups with distinct biological ancestry. Working collaboratively with local Indigenous groups in California and Mexico, we report genome-wide data from 87 ancient Californian individuals and 40 ancient Mexican individuals ranging from 7600 to 200 years old (BP), which we co-analyze with previously reported data. We find evidence for movement of ancestry related to ancient and present-day individuals from Northwest Mexico, who would have likely spoken Uto-Aztecan languages, arriving in southern and central California from at least 5,300 BP that ultimately reached its highest level in regions where Uto-Aztecan speakers currently reside. Ancient individuals from Baja California share more alleles with the earliest (5,200 BP) Central California individual than with later Indigenous Californians, potentially reflecting the first “Hokan” language associated populations in Central California having a high degree of relatedness with groups like those in Baja, and subsequent mixture transforming much of their ancestry. We find that ancient California individuals harbor increased affinity to the ~12,800 year old Anzick-1 individual associated with the Clovis culture compared to ancient northwest Mexico groups, which are on a different ancient lineage. Lastly, we find that some of the ancient groups from California and northwest Mexico had very small population sizes, similar to that of ancient Patagonian groups and significantly lower than ancient groups from the Andes and the Ceramic-period Caribbean.

S41. Populations evolving: Modeling genetic variation to understand evolutionary processes

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 260. 1,000 ancient genomes uncover 10,000 years of natural selection in Europe

Authors:

M. K. Le[1], O. Smith[1], A. Akbari[2,3], A. Harpak[1,4], D. Reich[2,3,5], V. M. Narasimhan[1]; [1]The Univ. of Texas at Austin, Austin, TX, [2]Harvard Med. Sch., Boston, MA, [3]Harvard Univ., Cambridge, MA, [4]Dell Med. Sch., Austin, TX, [5]Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract:

Ancient DNA has revolutionized our understanding of human population history, but its promise for changing our understanding of human genetic adaptation has not been realized to date, possibly due to limited sample sizes. We assembled genome-wide data from 1,291 individuals from Europe who lived in the past 10,000 years, increasing the total number of individuals for this type of analysis by ~6-fold. The dataset is large enough to not only detect selection, but also to resolve its timing into three major epochs—the Neolithic period, the Bronze Age and the Historical period—when changes in culture may have led to changes in selection pressures. We identified 26 genetic variants with changes in frequency inconsistent with random drift alone in one or more periods, suggestive of selection. A majority of these signals are previously undetected and 7 are in genes expressed on the surfaces of immune cells and are associated with host response to bacterial and viral pathogens. We also observe differences in functions associated with genes affected by selection in each epoch. Signals we find during the agricultural transition (Neolithic period) are associated with obesity, diet and lipid metabolism-related phenotypes. In this period, we also observed a signal at a locus that confers immunity to Salmonella infection at the same time as human adaptation to Salmonella is first observed. In the Bronze Age, selection signals were enriched near genes involved in pigmentation and immune-related traits, including at a key human protein interactor of SARS-CoV-2. It is only in the transition to the historical period that we observe most of the signals of selection that have been previously reported, as expected if the power of previous studies to detect signals was limited to the last few millennia. Finally, we detect signals of polygenic selection, including variants contributing to auto-immune and cardiovascular disease, and find that natural selection in the historical period may have increased the prevalence of multiple common diseases in contemporary Europeans compared to what would be expected if they carried the genomes of prehistoric Europeans.

S41. Populations evolving: Modeling genetic variation to understand evolutionary processes

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 261. Causal effects on complex traits are similar across local ancestries within admixed individuals

Authors:

K. Hou[1], Y. Ding[1], Z. Xu[1], Y. Wu[1], A. Bhattacharya[1], R. Mester[1], G. Belbin[2], D. Conti[3], B. Darst[4], M. Fornage[5], C. Gignoux[6], X. Guo[7], C. Haiman[3], E. Kenny[8], M. Kim[4], C. Kooperberg[4], L. Lange[9], A. Manichaikul[10], K. North[11], N. Nudelman[12], U. Peters[4], L. Rasmussen-Torvik[13], S. Rich[10], J. Rotter[14], H. Wheeler[15], Y. Zhou[4], S. Sankararaman[1], B. Pasaniuc[16]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Univ. of Southern California, Los Angeles, CA, [4]Fred Hutchinson Cancer Ctr., Seattle, WA, [5]Univ. of Texas Hlth.Sci. Ctr. at Houst, Houston, TX, [6]Univ. of Colorado Denver, Anschutz Med. Campus, Aurora, CO, [7]Harbor-UCLA Med. Ctr., Torrance, CA, [8]Icahn Sch. of Med. at Mt Sinai, New York, NY, [9]Univ. of Colorado Denver, Aurora, CO, [10]Univ. of Virginia, Charlottesville, VA, [11]Univ. of North Carolina, Chapel Hill, NC, [12]Rutgers Univ., Piscataway, NJ, [13]Northwestern Univ., Chicago, IL, [14]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [15]Loyola Univ. Chicago, Chicago, IL, [16]Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract:

Individuals of admixed ancestries (e.g., African Americans) inherit a mosaic of ancestry segments (local ancestry) originating from multiple continental ancestral populations. Their genomic diversity offers the unique opportunity of investigating genetic effects on disease across multiple ancestries within the same population. Quantifying the similarity in causal effects across local ancestries is paramount to studying genetic basis of diseases in admixed individuals (Hou et al. 2021 Nat Genet). Such similarity can be defined as the genetic correlation $R_{admix}=Cor[b_{afr},b_{eur}]$ of causal effects across African (b_{afr}) and European (b_{eur}) local ancestry backgrounds. Existing studies investigating causal effects variability across ancestries focused on cross-continental comparisons; however, such differences could be due to heterogeneities in the definition of environment/phenotype across continental ancestries. Studying genetic effects within admixed individuals avoids these confounding factors, because the genetic effects are compared across local ancestries within the same individuals.

Here, we introduce a new method that models polygenic architecture of complex traits to quantify R_{admix} across local ancestries. We model genome-wide causal effects that are allowed to vary by ancestry and estimate R_{admix} by inferring variance components of local ancestry-aware genetic relationship matrices. Our method is accurate and robust across a range of simulations. We analyze 38 complex traits in individuals of African and European admixed ancestries (N = 53K) from: Population Architecture using Genomics and Epidemiology (PAGE), UK Biobank and *All of Us*. We observe a high similarity in causal effects by ancestry in meta-analyses across traits, with estimated $R_{admix}=0.95$ (s.e. 0.01), much higher than correlation in causal effects across continental ancestries (0.85 (s.e. 0.01); Shi et al. 2021 Nat Commun). High estimated R_{admix} is also observed consistently for each individual trait.

We replicate the high correlation in causal effects using regression-based methods from marginal GWAS summary statistics. We also report realistic scenarios where regression-based methods yield inflated estimates of heterogeneity-by-ancestry due to local ancestry-specific tagging of causal variants, and/or polygenicity. Among regression-based methods, only Deming regression is robust enough for estimation of correlation in causal effects by ancestry. In summary, causal effects on complex traits are highly similar across local ancestries and motivate genetic analyses that assume minimal heterogeneity in causal effects by ancestry.

S41. Populations evolving: Modeling genetic variation to understand evolutionary processes

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 262. The cumulative effects of natural selection on protein truncating variants is associated with fecundity, mortality, morbidity, and life expectancy in humans

Authors:

H. Vy[1], D. M. Jordan[1], D. J. Balick[2], R. Do[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Harvard Med. Sch., Boston, MA

Abstract:

The accumulation of deleterious mutations leads to the reduction of an individual's fitness. However, this relationship is complex and could involve various indirect paths under different selective mechanisms. As the fitness of an individual is attributable to a few components (fertility, fecundity, mating success, etc.), the action of selection could be the integration of multiple selective episodes, each driven by a component of fitness. A recent study showed that fitness reduction due to the accumulation of heterozygous protein truncating variants (PTV), measured by aggregating the selection coefficients of genes affected by those variants, is associated with reduced reproductive success. In this study, we applied a similar measurement of fitness effect to further investigate the correlation between negative selection and direct components of fitness, including fecundity, morbidity and life expectancy. We estimated the reduction of fitness of an individual as delta(w)=1-exp(-sum($s_{het[i,g]}$)), where $s_{het[i,g]}$ is the pre-computed selection coefficients for gene g with PTVs from individual i. We calculated fitness reduction (delta(w)) in the exome sequences of 154,265 individuals from the UK Biobank and tested association with three fitness components: fecundity, morbidity and life expectancy. In univariate regression, we observed significant association of delta(w) with increased fecundity ($P = 2.05\text{x}10^{-3}$), morbidity ($P = 4.04.43\text{x}10^{-53}$ $4\text{x}10^{-3}$) and life expectancy ($P = 4.43\text{x}10^{-5}$). In multivariable regression, significant association of fitness reduction was observed for all three fitness components, suggesting independent contributions of each trait to fitness. Importantly, the relative variance explained was 27% for morbidity, 26% for life expectancy and 14% for fecundity. Our results provide genetic evidence that strong selection directly acts on these fitness-related traits. This further comprises the first quantitative estimation of the relative contributions of these traits to the net fitness of humans.

S41. Populations evolving: Modeling genetic variation to understand evolutionary processes

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 263. A mutation rate model at the basepair resolution identifies new mutagenic effects and improves population genetics inference

Authors:

D. Lee[1], V. Seplyarskiy[1], E. Koch[1], J. Lichtman[2], H. Luan[2], S. Sunyaev[3]; [1]Harvard Med. Sch., Boston, MA, [2]NGM Biopharmaceuticals, South San Francisco, CA, [3]Harvard Med. Sch./Brigham & Women's Hosp., Boston, MA

Abstract:

De novo mutations occur with substantially different rates depending on genomic location, sequence context and DNA strand. The success of many human genetics techniques, especially when applied to large population sequencing datasets with numerous recurrent mutations depends strongly on assumptions about the local mutation rate. Such techniques include estimation of selection intensity, inference of demographic history and mapping of rare disease genes. Here, we present Roulette, a mutation rate model at the basepair resolution. Roulette incorporates known determinants of local mutation rate and is shown to be more accurate than existing models, halving the amount of unexplained variance in synonymous variants compared to previous models. It also correctly predicts the fraction of tri-allelic SNVs. The analysis of significant deviations from the model predictions revealed a 10-fold increase of mutation rate in nearly all genes transcribed by Polymerase III suggesting a new mutagenic mechanism associated with transcription. We also detected an accelerated mutation rate within transcription factor binding sites actively utilized in testis, with no apparent effect of other binding sites. Incorporation of Roulette improves estimates of selective constraint measured as intolerance to protein truncating variants, and helps discriminate between models of human population history.

S42. The methylome and transcriptome of complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 266. Single-cell transcriptional hallmarks of Alzheimer's disease across 427 individuals

Authors:

Y. Tanigawa[1,2], N. Sun[1,2], W. F. Li[1,2], D. von Maydell[1,2], C. A. Boix[1,2], L. A. Akay[1,2], K. Galani[1,2], H. Mathys[1,2,3], D. A. Bennett[4], L-H. Tsai[1,2], M. Kellis[1,2]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [4]Rush Univ. Med. Ctr., Chicago, IL

Abstract:

Alzheimer's Disease (AD) is multifaceted, with many implicated biological pathways, across diverse cell types. The heterogeneous phenotypic manifestation, beyond the common characteristic signature of Amyloid beta plaque, across cognition, pathology, and treatment response is well recognized, but the molecular and cellular heterogeneity of AD remains uncharacterized at genomic and cellular resolution.

Here, we use single-cell RNA-seq profiling of 1.9 million cells from 430 human dorsolateral prefrontal cortex post-mortem brain samples across age-matched AD and non-AD individuals spanning all stages of AD progression. To assess heterogeneous molecular manifestation of the key biological processes, we develop a regularized multivariate differential expression analysis framework and identify 1621 gene expression patterns across 6 major cell types (implicating 1391 unique genes) associated with AD, which we cluster into 30 transcriptional hallmarks (Tx1-Tx30).

Our 30 transcriptional hallmarks capture several known cellular and pathological signatures in AD, pinpointing their candidate driver genes and cell types of action, and are associated with distinct phenotypic enrichments. For example, cytoplasmic translation in oligodendrocytes (Tx12) was most associated with early AD (p-value=5.3×10^{-7}) but not late AD ($p>0.9$) changes; cytoplasmic translation in oligodendrocyte precursor cells (Tx23) instead showed the strongest association ($p=1.0\times10^{-9}$) with amyloid level in the cortex; and response to zinc ion in astrocytes (Tx16) was most associated with neuritic plaque burden ($p=2.6\times10^{-9}$).

Using combinations of hallmark burdens, we classify our 430 donors into 12 AD and non-AD subtypes. Four and three of those groups were enriched in AD cases and non-AD individuals, respectively, and the other five were balanced between AD cases and controls. One of the AD groups driven by oligodendrocyte-associated hallmarks was preferentially enriched in neuritic plaque burden, which is more directly indicative of neuronal damage.

We found that genetic variants associated with single-cell gene expression (sc-eQTLs) across our 430 samples predict these transcriptional hallmarks with high accuracy (R=0.4), enabling us to both predict dysregulated AD hallmarks, risk groups, and subtypes across individuals decades before symptoms occur.

Overall, our results pave the way towards pathway-level therapeutic development, personalized prognosis and treatment tuning in AD, and more generally towards genetics-based prognosis, transcriptional subtyping, and clinical trial design in complex and heterogeneous traits even in inaccessible tissues.

S42. The methylome and transcriptome of complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 267. A cell-type-specific enhancer-gene map built from multimodal assay of RNA and ATAC-seq in 160,000 single cells pinpoints causal variants and genes in human diseases

Authors:

S. Sakaue[1], K. Weinand[1], K. Dey[2], K. Jagadeesh[3], A. McDavid[4], M. Kanai[3], Accelerating Medicines Partnership® RA/SLE Programand Network, L. T. Donlin[5], K. Wei[1], A. Price[2], S. Raychaudhuri[6]; [1]Harvard Med. Sch., Boston, MA, [2]Harvard Sch. of Publ. Hlth., Boston, MA, [3]Broad Inst. of MIT and Harvard, Allston, MA, [4]Univ. of Rochester, Rochester, NY, [5]Hosp. for Special Surgery Res. Inst., New York, NY, [6]Brigham and Women's Hosp., Brookline, MA

Abstract:

After 20 years of genome-wide association study, rarely have we identified causal variants or genes while defining causal mechanisms for each locus is essential to develop therapeutics. Statistical fine-mapping cannot narrow down causal variants in complete linkage. Causal genes are even harder to determine since most causal variants lie in non-coding regions, may regulate distant genes, and employ context specific regulatory mechanisms. Previous methods to link causal variants to genes have been based on bulk tissues (e.g. ABC and EpiMap), obscuring fine-scale regulation. To address this, we used single-cell multiome technology which jointly assays RNA and chromatin accessibility by ATAC-seq. We developed a new statistical method, SCENT, Single-Cell ENhancer Target gene map, to model association between chromatin accessibility and gene expression across individual single cells. SCENT uses a bootstrapping framework and shows well controlled type I error, despite the variable nature of gene counts in scRNAseq. We applied SCENT to 9 sc-multiome datasets (160565 cells, 8 cell types) and identified 99476 cell-type-specific enhancer-gene links. Unsurprisingly, all accessible regions were enriched in causal eQTL variants (PIP>0.2) by 2.7X than all cis-regions (GTEx). But strikingly, SCENT enhancers were more enriched in causal variants by 19.3X; this is much higher than those from conventional methods that link peaks to genes by correlation (e.g. ArchR, 2.5X). Further, SCENT enhancers were more enriched in causal GWAS variants in FinnGen (31.6X; 1046 traits) and UK Biobank (73.2X; 94 traits) than non-SCENT accessible regions with matched TSS distance, which is higher than bulk methods (ABC, 16.3X and EpiMap, 13.9X). On Mendelian diseases, SCENT enhancers harbored a larger number of ClinVar variants (1.9X) than distance matched regions, suggesting utility in interpreting both common and rare variants. SCENT enhancer-gene links were more likely to be experimentally validated by CRISPR-Flow FISH (4.5X, P=1.8x10^{-9}). We then sought to interpret GWAS by SCENT, which linked 4124 causal variants to genes. As examples, we (1) identified a single causal variant (rs72928038) in *BACH2* in rheumatoid arthritis (RA), validated by MPRA and base-editing, (2) discovered new variant-gene link in RA (rs59483192 to *RBPJ*) in T cells, supported by promotor-capture Hi-C, and (3) prioritized rs12592845 in height in Europeans regulating *FBN1* in fibroblasts, supported by an *FBN1* missense variant affecting height in Peruvians. Together, SCENT effectively defined causal variants and genes to parse the logic of GWAS loci and will contribute to a comprehensive enhancer-gene map.

S42. The methylome and transcriptome of complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 268. Combining comparative DNA methylation profiling and patient EHR data to characterize enhancer-gene networks in T2D genetic risk

Authors:

V. Agan[1], T. Scott[1], J. Sealock[1], A. X. Miranda[1], M. Brissova[1], E. Hodges[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Sch. of Med., Nashville, TN

Abstract:

Genome-wide association studies (GWAS) have shown that Type-2-diabetes (T2D)-associated single-nucleotide polymorphisms (SNPs) are predominantly noncoding and enriched within pancreatic islet-specific enhancers. Thus, most of the genetic susceptibility to T2D implicates perturbed islet transcription as a driving mechanism in impaired islet function. However, the regulatory effects exerted by these regions remain elusive. Our lab has shown that DNA hypomethylated regions (HMRs) identify putative cell-type specific enhancers and contain SNPs linked to cell-type relevant traits, including T2D. Furthermore, the genes targeted by these regions are highly enriched for cell-type specific biological processes. By intersecting epigenomic and genotype-phenotype data, we functionally profiled islet enhancer HMRs harboring variants associated with T2D and related traits. We generated DNA methylation profiles of islets from donors without T2D using whole genome bisulfite sequencing (WGBS) and curated a reference islet *hypo*-methylome, capturing over 35,000 HMRs across all donors. With this novel dataset, we compared DNA methylation patterns across islets and other cell types and tissues to identify HMRs unique to (n=4858) or shared with (n=30863) islets. Given a critical function of islets is to control blood glucose levels, we included HMR datasets of functionally related cell types to better delineate regions that are central to islet-specific gene activity. We found that genes associated with islet-specific HMRs are uniquely enriched for pathways regulating development and gene expression in β-cells (FDR $q < 0.05$), as opposed to shared islet HMRs. In agreement with our gene set analysis, between 44-47% of islet-specific HMRs contain binding motifs of key islet transcription factors -- FOXA2, NEUROD1, and PAX6 -- responsible for regulating insulin gene transcription and pancreatic development. Further supporting their distinct function, islet-specific HMRs are highly enriched for GWAS SNPs with T2D ($p = 2.3x10^{-10}$) and related traits, like glycated hemoglobin levels ($p = 6.5x10^{-7}$) and serum creatinine levels ($p = 2.62x10^{-6}$). Altogether, we show that putative enhancer islet-specific HMRs play a role in the gene regulatory networks of islets. Furthermore, these regions associate with pathways, that when disrupted, contribute to β-cell dysfunction. These studies improve our understanding of how changes in gene regulatory activities mediate T2D risk, which could be used for identification of therapeutic targets.

S42. The methylome and transcriptome of complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 269. Universal DNA methylation age across mammalian tissues

Authors:

A. Lu[1,2], Z. Fei[3], A. Haghani[1,2], K. Raj[4], S. Horvath[1,2,3]; [1]Altos Labs, San Diego, CA, [2]Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of California Los Angeles, Los angeles, CA, [3]Dept. of Biostatistics, Fielding Sch. of Publ. Hlth., Univ. of California Los Angeles, Los Angeles, CA, [4]Altos Labs, Cambridge, United Kingdom

Abstract:

ABSTRACT (1956 characters/2300 limit)Aging is often perceived as a degenerative process resulting from random accrual of cellular damage over time. In spite of this, age can be accurately estimated by epigenetic clocks based on DNA methylation profiles from almost any tissue of the body. Since such pan-tissue epigenetic clocks have been successfully developed for several different species, it is difficult to ignore the likelihood that a defined and shared mechanism underlies the aging process. To address this, we generated 11,754 methylation profiles of 59 tissue-types derived from 185 mammalian species, representing 19 taxonomic orders. Each profile is constituted by methylation measurements of 36,000 cytosines located in stretches of DNA that are highly-conserved across mammals. These samples ranged in age from prenatal to 139 years (bowhead whale). The considered species had maximum life span ranging from 1.9 (cinereus shrew) to 211 years (bowhead whale) and adult weight ranging from 0.004 (proboscis bat) to 100,000 kilograms (bowhead whale). From these, we identified and characterized specific cytosines, whose methylation levels change with age across mammalian species. These cytosines are greatly enriched in binding sites of polycomb repressive complex 2, which represses expression of genes involved in mammalian developmental process. Genes that are proximal to these CpGs are associated with age-related traits, including birth length, age at menarche, human longevity, leukocyte telomere length, Alzheimer's disease, etc. From this DNA methylation data set, we constructed two age predictors, each with a single mathematical formula, termed universal pan-mammalian clocks that are accurate in estimating ages ($r>0.96$) of any mammalian species and tissue. These universal clocks possess several features that are highly applicable to aging studies including the ability to predict mortality risks in human, evaluate mouse anti/pro-aging interventions, track processes in OSKM-based reprogramming of human cells, etc. Collectively, these new observations support the notion that aging is indeed evolutionarily conserved, deterministic and coupled to developmental processes across all mammalian species - a notion that was long debated without the benefit of this new and compelling evidence.

S42. The methylome and transcriptome of complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 270. Fine mapping of Alzheimer's disease susceptibility loci prioritizes functional variants within myeloid cell enhancers

Authors:

A. Ravi[1], K. de Paiva Lopes[1,2], J. Humphrey[1], R. Vialle[1,2], C-Y. Lee[3], K. Townsley[1], D. A. Knowles[4], K. Brennand[3,1], T. Raj[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Rush Univ. Med. Ctr., Chicago, IL, [3]Yale Sch. of Med., New Haven, CT, [4]Columbia Univ., New York, NY

Abstract:

Genome-wide association studies (GWAS) and quantitative trait locus (QTL) analysis have identified many common variants and genes associated with Alzheimer's Disease (AD) and highlighted the role of myeloid cells (microglia, monocytes, and macrophages) in the pathogenesis of AD. However, linkage disequilibrium (LD) often makes it challenging to discern true functional variants and their molecular mechanisms in AD. Our aims are to (i) prioritize common genetic variants that may indicate a causal effect; (ii) understand the role of these variants in myeloid cells; and (iii) perform functional validation of these variants. We perform statistical and functional fine mapping of 75 risk loci from the largest AD GWAS to date (Bellenguez et al. 2022), using PolyFun and SuSiE, to identify putative functional variants. We perform statistical enrichment analyses by integrating published Activity-by-Contact (ABC) model enhancer data from 132 human cell types. We also perform a random-effects meta-analysis of 10 myeloid-cell expression-QTL (eQTL) datasets (**meta-myeloid:** 1,763 samples, 1,187 unique donors), followed by colocalization analysis to identify significant myeloid-cell variants and genes in AD. Our fine mapping results reveal 60/75 AD risk loci have at least one significant SNP (posterior inclusion probability (PIP) > 0.1). While several loci contain missense variants, particularly in the *TREM2*, *SHARPIN*, *ABI3*, *PLCG2,* and *MME* loci, most SNPs identified lie in noncoding regions, requiring further validation. Preliminary analyses show significant enrichment in myeloid cell enhancers identified by the ABC model, with the strongest enrichment seen in microglia (enrichment ratio = 21.2, $p = 2x10^{-12}$). One variant, rs4714447, in the *TREM2* locus (PIP = 0.28), is found across all myeloid cell enhancers, and is significantly associated ($p = 3x10^{-22}$) with *TREM1* gene expression in the meta-myeloid eQTL. Colocalization analysis with the meta-myeloid eQTL identifies 43 significant (H4 posterior probability > 0.5) GWAS loci-eQTL gene pairs. These findings support the hypothesis that dysregulation of myeloid cells is an important factor in AD risk. We are currently conducting a massively-parallel reporter assay (MPRA) of 11,550 fine-mapped and eQTL variants in iPSC microglia, testing whether these variants act as enhancers using a reporter construct. Overall, these results illustrate the significance of myeloid cells in AD risk and provide a comprehensive list of regulatory variants and genes that may serve as therapeutic targets.

S42. The methylome and transcriptome of complex traits

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 271. Molecular mechanisms of healthy blood aging at single-cell resolution

Authors:

E. Bader[1,2], M-J. Fave[1], M. Agbessi[1], E. Gbeha[1], J. Uzinovic[1], D. Soave[3], K. Skead[1,2], V. Bruat[1], P. Awadalla[1]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Dept of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]Wilfrid Laurier Univ., Waterloo, ON, Canada

Abstract:

All individuals experience aging of the immune system, yet some elderly individuals maintain exceptional immune function. The direct effects of aging on immune cell function have been difficult to dissect because of the variability across individuals conferred by genetic and environmental factors. From a cohort of 230,000 Canadians, we show how blood cell profiles can predict healthy immune cell aging and risk for cancer. We selected 400 individuals from the extremities of the age and immune health spectrums and performed single cell RNA sequencing on over 500 000 cells, together with whole genome sequencing and chromatin profiling on bulk CD45+ cells to reveal genetic, epigenetic, and transcriptional factors underlying healthy immune aging. We observe a significant association between an increase in the proportion of CD4+ T memory cells and reduced immune health in females. Across 22 cell populations, 245 genes show sex and cell-type specific differential gene expression associated with immune health, and are enriched in pathways for splicing (*SLC38A2*), iron homeostasis (*FTH1, FTL)*, and immune signalling (*KLF6, DUSP1, CD52*). These results suggest that immune cell function rather than composition contributes to the variation in immune health as one ages, and that canonical age-related pathways are exacerbated in individuals with unhealthy immune profiles. Most genes differentially expressed with immune health are expressed at similar levels in healthy aged and young individuals despite a 20+ year age difference. Thus, maintenance of gene expression at these loci over time is one mechanism that contributes to healthy immune cell aging. Heritable and epigenomic contributions to variation in immune health were captured following context-dependent chromatin-accessibility QTLs (caQTL) in bulk samples and expression QTLs (eQTLs) across immune cell types. We identified 1105 context-dependent eQTLs enriched in innate cell types, suggesting their importance in healthy immune aging. In adaptive cell types, eQTLs associated with healthy aging are enriched in polycomb repressed and quiescent chromatin regions, suggesting a possible mechanism for regulating gene expression at these loci. To identify regulatory mechanisms on which context-dependent eQTLs might be acting, we performed cell type specific colocalization analyses. We identified 12 and 7 colocalization events with canonical ca-QTL and context specific ca-QTLs, respectively. Collectively, our results demonstrate how natural variation in healthy agers can help to uncover mechanisms that prevent or protect dysregulated immune function during aging.

S43. Using omics to dissect GWAS signals

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 274. Proteogenomic analysis in UK Biobank identifies potential proteomic consequences of genetic susceptibility to Parkinson's disease

Authors:

C. Robins[1], P. Gormley[2], J. Davitte[1], P. Surendran[3], E. Arciero[3], J. Todd[4], C-F. Xu[3], R. Scott[3]; [1]GlaxoSmithKline, Collegeville, PA, [2]GlaxoSmithKline, Cambridge, MA, [3]GlaxoSmithKline, Stevenage, United Kingdom, [4]Univ. of Oxford, Oxford, United Kingdom

Abstract:

Parkinson's Disease (PD) is a devastating neurodegenerative disease with huge unmet medical need for disease modifying and preventative therapies. To identify plasma proteins potentially linked to PD susceptibility, we calculated and tested the association of PD polygenic risk scores (PRS) with the abundance of 1,462 proteins using genetic and proteomic data from 29,963 European participants in UK Biobank without a prevalent diagnosis of PD. The PD PRS comprised 140 independent risk variants and was derived by applying LD-clumping and P-value thresholding (250 Kb windows, $R2 > 0.1$, $p < 1x10\text{-}5$) to publicly available GWAS summary statistics of PD after excluding UKB samples. The PRS was associated with the abundance of 14 proteins at a Bonferroni-adjusted significance threshold ($p < 3.4x10\text{-}5$). Of these 14 proteins, five have cis-protein quantitative trait loci (cis-pQTL, within 1 Mb of the protein-coding gene) that were included in the PD PRS. For each of these five proteins, we investigated the influence of cis-pQTLs by re-calculating the PRS after cis-pQTL exclusion and re-testing for associations with protein abundance. Two proteins, encoded by *ADAM15* and *BST1*, had no significant association with the PD PRS when their cis-pQTLs were excluded. Thus, most of the identified associations (12/14) were not immediately explained by cis-pQTLs and may be polygenic in nature. To test if the proteins identified had a causal role in disease or were a consequence of disease pathophysiology, we performed bi-directional two-sample Mendelian Randomization (MR) analyses. These analyses compared pQTL data from >37k European UK Biobank samples to PD GWAS summary statistics, and vice versa, using the inverse-variance weighted method. Preliminary MR results implicated none of the tested proteins to have a causal impact on PD, but suggested that three proteins, encoded by *CDH2*, *GGH*, and *SCARB2*, had abundance changes that are potential consequences of genetic susceptibility to PD (false discovery rate < 0.05).

S43. Using omics to dissect GWAS signals

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 275. Identifying potential causal genes by integrating molecular intermediate phenotypes and GWAS signals in psychiatric disorders

Authors:

H. Gedik, T. Nguyen, S-A. Bacanu; Virginia Commonwealth Univ., Richmond, VA

Abstract:

Background: Psychiatric disorders (PDs) have complex etiology that is not elucidated by GWAS alone. Post GWAS methods allow us to infer association between PDs and molecular intermediate phenotypes, e.g., transcript and protein abundance. Applying such tools, e.g., Transcriptome-wide Association studies (TWASs) and Proteome-wide Association studies (PWASs), to PDs GWAS was shown to increase genetic signal detection. In this research, we aim to identify the risk loci mediating molecular intermediate phenotypes, in blood and brain tissues, between genetic variants and PD diagnosis using the latest i) GWAS of PDs and ii) transcriptomic and proteomic reference panels for a) brain and b) blood tissues.
Methods: We use the Summary Mendelian Randomization (SMR) tool to infer the association between gene expression/protein abundance and nine PDs in blood serum and brain tissues. For the blood serum, we employ the largest reference expression/protein-QTL (e/p-QTL) panels: i) eQTLGen (N=~31K) for blood TWAS ii) deCODE (N=~35K) for blood PWAS. Similarly, for brain tissues, we use the largest two available panels: i) Brain eMeta version 2 (N=~2K) from Qi et al. (SMR group) for brain TWAS and ii) ROS/MAP (N=~7k) Robins et al. for brain PWAS. Subsequently, to increase detection power of the less powered brain signals, we test brain blood concordance using a Bayesian paradigm.
Results: We provide Miami plots (brain - blood) for nine PDs. We find concordant brain - blood signals for cannabis use disorder in TWAS (*HYAL3 and NAA80*), for PTSD in TWAS and PWAS (*CCBL2*), SCZ TWAS (*ZNF823, RERE, NAGA, PCCB*) and BIP TWAS (*LMAN2L, GNL3, PLEC, ADD, TMEM258*). Similarly, we observe brain-blood concordance of CCBL2 for PTSD TWASs and PWASs. The only significant PWAS signal in ADHD is for TIE1 in blood PWAS, which is close to ADHD TWAS signal (MED8). Three disorders (BIP, SCZ and MDD) have common blood (*BTN2A2, BTN3A2, C4A, C4B, IER3, ZSCAN16 and VARS2*) and brain (*BAG6, MICA, TRDN and VWA7*) TWAS significant gene signals in HLA region. In SCZ and MDD, *BTN3A3 and MICB* were the shared blood PWAS signals in the HLA region. The leading blood AUD PWAS signal is *ADH1B*, which is biologically significant.
Conclusions: While most GWASs loci lack a clear biological explanation, TWAS/PWAS paradigms allow researchers to uncover genes that are etiologically related to traits. For instance, inflammation seems to be involved in many PDs, as suggested by the three well-powered diseases (SCZ, BIP and MDD) having overlapping signals in MHC region. Identification of such novel molecular pathways and protein targets help researchers to uncover avenues for therapeutic intervention and, possibly, clinical diagnosis in PDs.

S43. Using omics to dissect GWAS signals

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 276. Deconvolution of bulk RNA-seq reveals cell-type specificity mechanism in Alzheimer's disease

Authors:

J. Wen[1], Q. Sun[1], G. Li[1], X. Yang[2], J. Chen[1], Y. Shen[3], Y. Li[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ. of California, San Francisco, San Francisco, CA, [3]Inst. for Human Genetics, San Francisco, CA

Abstract:

Gene expression in the human brain is usually measured from tissue samples by RNA-sequencing, commonly referred to as bulk RNA-seq data. Though such bulk tissue transcriptomic profiles can reflect the etiology of brain-related diseases such as Alzheimer's disease (AD) to a certain degree, they cannot reflect the functional heterogeneity across cell types due to the mixture of multiple cell types. One gold-standard way to obtain cell-type-specific (CTS) gene expression is from single-cell (sc) or single-nuclei (sn) RNA-seq data, but they are cost-prohibitive for population samples, and also tend to be noisier than bulk RNA-seq data. Fortunately, with bulk RNA-seq data collected in study samples and increasingly rich scRNA-seq or snRNA-seq data accumulated in the public domain, the underlying cell type information could be inferred using computational methods. We here propose a new aggressive-sense cell type deconvolution method (EPIC-unmix), a two-step empirical Bayes method that integrates sc or snRNA-seq reference data and bulk RNA-seq data from target samples to update prior, improving accuracy of CTS expression inference in target samples. EPIC-unmix first infers CTS expression using reference as a prior, and then adds another layer of Bayesian inference based on the CTS expression inferred for the target samples. We applied EPIC-unmix to deconvolute ROSMAP bulk RNA-seq data from prefrontal cortex samples, leading to the estimation of CTS gene expression profiles specific to each ROSMAP sample. As it is impossible to accurately estimate CTS expressions for every gene in every cell type, it is important to prioritize genes that can be well inferred, separately in each cell type. To obtain such lists of high-confidence genes, we developed a gene selection strategy that combines multiple sources of brain sc and snRNA-seq data, as well as CTS inference from real data-based simulations. Downstream analysis of CTS gene expression, including identification of CTS differentially expressed (DE) genes, CTS eQTL analysis and functional annotations in the corresponding cell type for DE genes, suggested *IKZF1* as an AD risk gene functioning in microglia. Specifically, *IKZF1* is upregulated among AD patients (microglia DE p-value = 0.04) with three eQTLs residing in a candidate *cis* regulatory region (cCRE), all specific in microglia. Finally, we carried out CRISPRi experiment perturbing the cCRE in human pluripotent stem cell (hPSC)-derived microglia, which validated the regulatory role of the cCRE on *IKZF1* expression (p = 0.007 and 0.013 for two replicates). We believe that our EPIC-unmix provides a new paradigm to generate CTS mechanistic hypotheses.

S43. Using omics to dissect GWAS signals

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 277. Large-scale transcriptome-wide association study identifies novel candidate genes underlying prostate-specific antigen levels

Authors:

D. Chen[1], R. Dong[2], L. Kachuri[2], Y. Jiang[1], S. Berndt[3], J. Shelley[4], K. Schaffer[5], M. Machiela[3], N. D. Freedman[3], W-Y. Huang[3], S. Li[3], H. Lilja[6], S. Van Den Eeden[7], S. Chanock[3], C. Haiman[8], D. Conti[8], R. Klein[9], J. Mosley[5], T. Hoffmann[1], J. Witte[2], R. Graff[1]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Stanford Univ., Stanford, CA, [3]Natl. Cancer Inst., Rockville, MD, [4]Vanderbilt Univ. Sch. of Med., Nashville, TN, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Mem. Sloan Kettering Cancer Ctr., New York, NY, [7]Kaiser Permanente Northern California, Oakland, CA, [8]Univ. of Southern California, Los Angeles, CA, [9]Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Background: Prostate-specific antigen (PSA) is a serine protease produced by prostate epithelial cells that is used a serum biomarker for prostate cancer (PCa) detection. Deciphering the genetic basis of PSA levels may improve their utility for PCa screening. We conducted the first ever transcriptome-wide association study (TWAS) of PSA levels using genome-wide summary statistics from 95,768 PCa-free men of predominantly (90%) European ancestry. **Methods:** We used MetaXcan with gene expression multivariate adaptive shrinkage prediction models trained on Genotype-Tissue Expression (GTEx) tissue-specific expression quantitative trait loci (eQTL) v8 data. Tissue-specific analyses were conducted in whole blood and prostate tissue. Cross-tissue analyses used 49 distinct GTEx tissues. To assess novelty of the results, we evaluated whether known genome-wide significant variants were independent from the eQTLs predicting transcriptome-wide significant genes. **Results:** Our TWAS analyses identified 41 genes in whole blood, 39 genes in prostate tissue, and 153 genes across tissues that were significantly associated with PSA levels after Bonferroni correction. Of these, 17 genes were located on 19q13, which contains the kallikrein (*KLK*) gene family that modulates PSA synthesis and homeostasis. Assessment of independence relative to known GWAS variants yielded 8 novel genes: *CCNA2* (4q27)*, EXOSC9* (4q27)*, LRRC41* (1p34.1), *GPBP1L1* (1p34.1)*, TMEM69* (1p34.1) in whole blood and *SCUBE2* (11p15.4)*, PENK* (8q12.1), and *HEXIM1* (17q21.31) in prostate tissue. Of particular interest, *CCNA2* encodes the cell cycle regulator protein Cyclin-A2 that mediates G1/S phase transition and entry into mitosis. D-type cyclin proteins have been associated with androgen signaling, a pathway that regulates serum PSA levels. Cyclin 2A specifically, however, has not been previously associated with PCa, PSA levels, or androgen signaling. **Conclusions:** In this first TWAS examining genes associated with PSA levels in PCa-free men, we identified 8 novel genes associated with PSA levels. These findings yield novel hypotheses for genetic factors underlying PSA levels that should be further explored toward improving our understanding of the biology of PSA levels, with the ultimate goal of enhancing PSA screening.

S43. Using omics to dissect GWAS signals

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 278. Applications of Individual-level Imputed Transcriptomes in the UK Biobank

Authors:

X. Wang[1,2], C. Selvaggi[1], L-C. Weng[1,2], S. Choi[2,3], T. Cao[1,4], A. Jha[1,4], S. Koyama[2], P. Ellinor[1,2], A. Foulkes[1,4,5], S. Lubitz[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst., Cambridge, MA, [3]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [5]Dept. of Med., Harvard Med. Sch., Boston, MA

Abstract:

Gene expression is related to phenotypic variability. However, since tissue-specific gene expression is difficult to measure at scale, the relations between expression profiles and phenotypic variability can be difficult to infer. We applied a joint-tissue imputation approach to predict the *cis* genetic component of gene expression for 486,452 participants in the UK Biobank. We then conducted analyses to demonstrate the potential applications of individual-level transcriptome imputation. First, we performed transcriptome-wide association studies (TWAS) for six phenotypes (atrial fibrillation, coronary heart disease, type 2 diabetes [T2D], low-density lipoprotein cholesterol [LDL], glycated hemoglobin, electrocardiographic heart rate corrected QT interval) in their corresponding biologically relevant tissues. Second, we constructed individual-level imputed transcriptome risk scores (ITRS) using associated genes observed in the TWAS. We identified 1,082 significant gene-tissue-phenotype trios (e.g., *ANGPTL3* expression in the liver was positively associated with LDL) after Bonferroni correction ($P < 0.05/110,109$ tests), of which 85 (7.9%) are existing potential therapeutic targets. We observed significant associations between each ITRS and the phenotypes in an independent sample, with hazards ratios (HRs) ranging from 1.30-1.39 per standard deviation (SD) increase of the ITRS for disease phenotypes and beta estimates ranging from 0.15-4.98 per SD increase of the ITRS for quantitative traits (all $P < 0.001$). We observed persistent associations but attenuated effect sizes with HRs ranging from 1.09-1.15 and betas ranging from 0.04-2.39 (all $P < 0.001$) when we created ITRS from the subset of associated genes that were potential therapeutic targets. Our study provides evidence supporting the utility of individual-level imputed transcriptomes for the discovery of novel gene-phenotype associations, expression-based prediction of phenotypic risk, and identification of individuals with therapeutically targetable predisposition to disease.

S43. Using omics to dissect GWAS signals

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 279. A multi-ancestry genome-wide meta-analysis, fine-mapping, and target gene prioritization to characterize the genetic architecture of adiponectin

Authors:

V. Sarsani[1], S. Brotman[2], C. Spracklen[1]; [1]Univ. of Massachusetts, Amherst, MA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Previous genome-wide association studies (GWAS) for circulating adiponectin levels, a complex trait linked to type 2 diabetes and obesity, identified >20 loci. However, most of the findings were based only on populations of European ancestry and many of the target genes underlying the associations remain unknown. We conducted the largest multi-ancestry adiponectin GWA meta-analysis to date, in up to 46,434 individuals from the METSIM cohort and the ADIPOGen and AGEN consortiums. Study-specific association summary statistics were combined using a fixed effects, inverse variance-weighted approach implemented in METASOFT; heterogeneity statistics from the RE2 model and Bayes factors from MR-MEGA were also considered. We identified 22 loci associated with adiponectin ($P < 5x10^{-8}$), including 15 known and 7 novel loci. Among individuals of European ancestry, GCTA-COJO identified 14 additional secondary signals at the ADIPOQ, CDH13, HCAR1, and ZNF664 loci. Leveraging the multi-ancestry data, FINEMAP + SuSiE identified 21 causal variants (PP>0.9). Target genes at primary and secondary loci were prioritized using colocalization with gene expression (eQTLs), Summary Mendelian Randomization, and Expression Modifier Scores. Prioritized target genes of novel loci in adipose and related tissues include, among others: CSF1, RGS17, PDE3B, CTDNEP1, and SLC2A4 (formerly GLUT4). Identified through consensus of several colocalization approaches is CSF1, macrophage colony stimulating factor, which is known to increase adipose tissue growth and proliferation, which may, in turn, affect circulating levels of adiponectin. We applied an integrated approach for pathway enrichment (MAGMA, PoPS, VEGAS2, and Downstreamer) at novel loci, which identified SLC2A4, a glucose transporter involved in adipogenesis, type 2 diabetes, and insulin response pathways. Our findings provide new insights into the genetic architecture of adiponectin and further highlight both the importance of conducting genetic studies across multiple ancestries and integrating multiple methodologies for robust prioritizing of target genes.

S44. Voices from the community

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 282. The Silent Genomes 'Precision Diagnosis Study': A strategy to improve access to translational genomics research for Indigenous Peoples in Canada

Authors:

K. Jacob[1], S. McIntosh[1], B. Morgan[1], N. Caron[2], W. Wasserman[1], M. Tarailo-Graovac[3], L. Badalato[4], F. Bernier[3], K. Boycott[5], O. Caluseriu[6], L. Chad[7], G. Costain[7], K. Davis[8], I. De Bie[9], S. Dyack[10], C. R. Greenberg[11], A-M. Laberge[12], F. Lévesque[8], M. Nowaczyk[13], J-B. Rivière[9], V. Siu[14], L. Turner[15], A. Vandersteen[10], L. Arbour[1], A. Lehman[1], Silent Genomes Precision Diagnosis Consortium; [1]Univ. of British Columbia, Vancouver, BC, Canada, [2]Univ. of Northern British Columbia, Prince George, BC, Canada, [3]Univ. of Calgary, Calgary, AB, Canada, [4]Queen's Univ., Kingston, ON, Canada, [5]Univ. of Ottawa, Ottawa, ON, Canada, [6]Univ. of Alberta, Edmonton, AB, Canada, [7]Univ. of Toronto, Toronto, ON, Canada, [8]Univ. of Saskatchewan, Saskatoon, SK, Canada, [9]McGill Univ., Montreal, QC, Canada, [10]Dalhousie Univ., Halifax, NS, Canada, [11]Univ. of Manitoba, Winnipeg, MB, Canada, [12]Université de Montréal, Montreal, QC, Canada, [13]McMaster Univ., Hamilton, ON, Canada, [14]Western Univ., London, ON, Canada, [15]Mem. Univ., St. John's, NL, Canada

Abstract:

Even in the face of resilience and increasing self-determination, Indigenous Peoples continue to face healthcare inequities, including barriers to accessing genomic testing and underrepresentation in genomics research. The 'Precision Diagnosis Study' within the Silent Genomes Project aims to bridge this genomics divide by offering genomic sequencing for diagnostic purposes, solely to patients who self-identify as Indigenous and have undiagnosed rare disease.

Eligibility barriers were minimized to maximize potential benefit for as many patients as possible. Trio enrollments were not required, and multiple enrollment sites across Canada facilitated comprehensive geographic coverage. Team members were provided Indigenous cultural safety training with the aim of improving participant experience and quality of care. Multimodal education and consenting were implemented.

This project faced many challenges. Although Canadian research ethics guidelines provide a template for carrying out participatory research within Indigenous communities, there are no guidelines for Indigenous patient-based research outside of the community context. This project's focus on rare disease diagnosis only for Indigenous individuals presented a novel circumstance to ethics boards across Canada, contributing to lengthy delays in achieving ethics approvals, and in some cases approval was never granted. Other challenges included difficulty obtaining samples, overextended personnel, and delays with legal contracting between institutions.

Four years in, and nearing the end of the project, 11/15 enrollment sites across Canada are open for recruitment. Two sites have not yet achieved ethics approval, and 2 more had to withdraw due to limited capacity. Eighty families or singletons have been enrolled, meeting 40% of the enrollment target. Time from referral to sample procurement averaged 8.7 months, reflecting challenges for both participants and personnel. This project prompted 28 team members across Canada to complete Indigenous cultural safety training for the first time.

Our attempt to bridge the genomics gap in Canada has highlighted the need for policy updates in order to achieve success in the pursuit of health equity. There is clear need for guidelines on how to best carry out research involving Indigenous individuals when outside the realm of community-based research, as well as the need for a harmonized Indigenous research ethics review process for nation-wide projects. Without these changes, projects such as this may continue to struggle and inequities in genomic diagnosis will propagate.

S44. Voices from the community

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 283. Assessing Vietnamese American patient views toward incorporating genomics in primary care: A community engaged research approach

Authors:

A. Lemke[1], S. Choi[2], V. Dang[3], T. Dang[4], J-H. Yu[5]; [1]Norton Children's Res. Inst., Affiliated with the Univ. of Louisville Sch. of Med., Louisville, KY, [2]NorthShore Univ. Hlth.System, Evanston, IL, [3]Summit Sch. Inc., Elgin, IL, [4]Swedish Med. Group, Part of NorthShore, Chicago, IL, [5]Univ. of Washington Sch. of Med., Seattle, WA

Abstract:

Background: Achieving health equity in precision medicine remains a critical challenge due to the continued underrepresentation of diverse patient populations in research and barriers to genetic services. The goal of this community engaged research (CEnR) study was to explore Vietnamese American (VA) patient views toward incorporating genetic testing into their primary care experience, to better serve the local VA community within a healthcare system implementing population genetic screening. **Methods:** We conducted semi-structured interviews in preferred languages with primary care patients who self-identified as Vietnamese or Vietnamese American and employed rapid qualitative analysis (RQA) to identify key concepts. Consistent with the CEnR approach of community involvement in the research, community team members were trained in research methods and participated in study design, data collection, RQA, and reporting. **Results:** Twenty-two individuals participated and 13 (59%) completed interviews in Vietnamese. All but one participant reported their country-of-origin as Vietnam, arriving in the US between 1970-2021. Half of study participants were male and 16 (73%) were between the ages of 18 and 39 (range 18-70+). Fear was an important theme both related to and beyond genetics. More than a quarter of participants (27%) expressed fear of violence due to race. Participant-perceived challenges to genetic testing included lack of information, fear of results impact, cost, and privacy concerns. Participants suggested ways to overcome these barriers including reducing test cost, receiving information from a trusted physician, using preferred community education strategies, and convenient access to testing. Patients shared trusted sources from whom they would seek advice on genetic testing including a primary care doctor, other healthcare providers, family and friends, and reputable sources on the internet. **Discussion:** This study with VAs identified barriers, facilitators and messengers to genetic screening in a local healthcare context, and demonstrated how CEnR, coupled with RQA, is a promising approach for healthcare institutions as they identify needs and tailor strategies for implementing population genetic screening among patients from ethno-geographic communities. Further, strategies for VAs and other ethnic minority communities must contend with broader social issues such as fear of racial violence as a structural barrier to healthcare and public health. Future research is needed to understand the perspectives of ethno-geographic local communities to equitably realize benefits of population genetic screening.

S44. Voices from the community

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 284. The perspective of community gatekeepers on genomic risk information in the context of orofacial cleft in an African population

Authors:

A. Oladayo[1], O. Odukoya[2], V. Sule[1], I. Molobe[3], T. Busch[1], W. Adeyemo[3], L. Gowans[4], M. Eshete[5], W. Awotoye[1], A. Alade[1], B. Akodu[1], A. Adeyemo[6], P. Mossey[7], A. Prince[1], J. Murray[1], A. Butali[1]; [1]Univ. of Iowa, Iowa, IA, [2]Dept. of Community Hlth.and Primary Care, Coll. of Med., Univ. of Lagos, Lagos, Nigeria, [3]Univ. of Lagos/ Lagos Univ. Teaching Hosp., Lagos, Nigeria, [4]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [5]Addis Ababa Univ., Addis Ababa, Ethiopia, [6]NIH, Bethesda, MD, [7]Univ. of Dundee, Dundee, United Kingdom

Abstract:

Background: Little is known about the role of the community and faith organizations in decisions to undergo genomic testing/research and subsequent result management in the African population, particularly in the context of orofacial cleft (OFC). Understanding the opinions of religious and community gatekeepers who are influential members of the community and health system could inform the design of community engagement strategies and increase the exposure to the benefits of genomic testing/research, especially for Africans underrepresented in genomic research.
Aim: To explore the perspectives of religious and community gatekeepers concerning genomic risk information in the presence of OFCs in minority populations.
Methods: Using a validated semi-structured questionnaire, two trained moderators collected information on participants' views on genomic risk information (GRI) and OFC experience via twenty-five focus group discussions (FGDs) with 214 gatekeepers (community, ethnic, religious leaders, and traditional birth attendants) in Lagos, Nigeria from October to December 2021. Transcripts were generated from the audio recordings obtained from the FGDs. The codes from the thematic analysis were obtained using the deductive-inductive approach. The analysis was done manually and cross-referenced using NVivo to ensure rigor and reproducibility of the results.
Results: Three main themes that emerged from exploring the perspective of gatekeepers about GRI in this cohort are: knowledge, beliefs, and willingness to take action. Furthermore, mixed opinions were observed. While some participants believed that GRI could play a role in the diagnosis, management, and prevention of diseases, others expressed dissatisfaction to the outright rejection of such information. Also, religious and cultural beliefs were crucial to determining participants' understanding of OFCs and the acceptance and utilization of genomic risk information. Recommendations to help the public better understand and utilize GRI include promoting a synergistic relationship between community health and spiritual providers.
Conclusion: The results highlight the invaluable role that gatekeeper engagement could play in how their members respond to and act upon genomic risk information. However, the cultural and religious diversity across the African continent could lead to a variation in the population perception of genomic risk information, thus affecting generalizability. Further research which captures the trans-continental socio-cultural uniqueness is needed to capture the opinions that better represent the diversity of the African population.

S44. Voices from the community

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 285. Participants as partners: how research participants guide the approach for returning genetic health-related information in the All of Us Research Program

Authors:

C. Neben[1], S. Okumura[1], A. Boateng-Kuffour[1], L. Servais[1], S. Wallace[2], S. Wallace[1], W. McKennon[1], S. Pimental[1], S. Law[1], A. Chua[1], D. Lee[1], K. Gerwig[1], H. Lilligren[1], S. Topper[1], A. Zhou[1], All of Us Research Program; [1]Color Hlth., Burlingame, CA, [2]Seattle Children's Hosp., Seattle, WA

Abstract:

Background: In late 2022, the NIH All of Us Research Program will begin returning health-related genetic results about hereditary disease risk (HDR) and pharmacogenomics (PGx) to participants. The cohort is unique as it reflects the demographic diversity of the US, with an emphasis on individuals who have been historically underrepresented in biomedical research. This, along with the program’s direct-to-participant return of results model, compelled the program to develop HDR and PGx reports that would be understood across a diverse cohort. Methods: We first identified genetic knowledge, self-efficacy, and risk numeracy as the set of key concepts that should be in the three report types the program will return (Positive and Uninformative (i.e., Negative) HDR reports and a PGx report). We then conducted three rounds of qualitative interviews in a genetic testing-naive population (n=21) to guide report development. With each round, we also iterated on the survey to assess report comprehension. After concept saturation (>90% pass rate), we administered the survey online to another set of participants to quantitatively evaluate their understanding of the revised Positive HDR (n=342), Uninformative HDR (n=278), and PGx (n=205) reports.Results: In the qualitative phase, mean comprehension of the Positive HDR, Uninformative HDR, and PGx reports improved by 18.0%, 20.0%, and 33.3%, respectively. Across all reports, participants had slightly higher comprehension of self-efficacy compared to genetic knowledge. Risk numeracy comprehension, which was only included in the Positive HDR report, was low (66.0%), and these concepts were eliminated after the first round. We found that report format, key concept location, and scientific jargon all impacted participant understanding. For example, participants struggled with the terms “health care provider”, “DNA change”, and “pharmacogenomics", which were widely used in earlier report versions. In the quantitative phase, about half of participants were 45+ years of age (48.6%), female (63.8%), and non-white (48.2%); had an associate degree or less education (50.5%); and earned <$75,000 annually (60.4%). Mean report comprehension was extremely high (Positive HDR, 97.1%; Uninformative HDR, 96.6%; and PGx, 98.1%), and self-efficacy was again slightly higher than genetic knowledge. Conclusion: These results demonstrate that an iterative approach to genetic test report development that incorporates participant feedback can maximize participant comprehension and should be considered by other research programs to responsibly return health-related genetic information to large, diverse cohorts.

S44. Voices from the community

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 286. Effectiveness of the *Family Heart Talk* communication tool in improving family member screening for dilated cardiomyopathy: Results of a randomized trial

Authors:

D. D. Kinnamon[1], E. Jordan[1], G. J. Haas[1], M. Hofmeyer[2], E. Kransdorf[3], G. A. Ewald[4], A. A. Morris[5], A. Owens[6], B. Lowes[7], D. Stoller[7], W. H. W. Tang[8], S. Garg[9], B. H. Trachtenberg[10], P. Shah[11], S. V. Pamboukian[12], N. K. Sweitzer[13], M. T. Wheeler[14], J. E. Wilcox[15], S. Katz[16], S. Pan[17], J. Jimenez[18], K. D. Aaronson[19], D. P. Fishbein[20], F. Smart[21], J. Wang[22], S. S. Gottlieb[23], D. P. Judge[24], C. K. Moore[25], J. O. Mead[1], D. J. Bowen[20], G. S. Huggins[26], H. Ni[1], W. Burke[20], R. E. Hershberger[1], DCM Precision Medicine Study of the DCM Consortium; [1]The Ohio State Univ., Columbus, OH, [2]Medstar Res. Inst., Washington Hosp. Ctr., Washington, DC, [3]Smidt Heart Inst., Cedars-Sinai Med. Ctr., Los Angeles, CA, [4]Washington Univ., St. Louis, MO, [5]Emory Univ. Sch. of Med., Atlanta, GA, [6]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [7]Univ. of Nebraska Med. Ctr., Omaha, NE, [8]Heart, Vascular, and Thoracic Inst., Cleveland Clinic, Cleveland, OH, [9]Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [10]Houston Methodist DeBakey Heart and Vascular Ctr., J.C. Walter Jr. Transplant Ctr., Houston, TX, [11]Inova Heart and Vascular Inst., Falls Church, VA, [12]Univ. of Alabama at Birmingham, Birmingham, AL, [13]Sarver Heart Ctr., Univ. of Arizona, Tucson, AZ, [14]Stanford Univ. Sch. of Med., Palo Alto, CA, [15]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [16]New York Univ. Langone Med. Ctr., New York, NY, [17]Westchester Med. Ctr. & New York Med. Coll., Valhalla, NY, [18]Miami Cardiac & Vascular Inst., Baptist Hlth.South, Miami, FL, [19]Univ. of Michigan Med. Ctr., Ann Arbor, MI, [20]Univ. of Washington, Seattle, WA, [21]Louisiana State Univ. Hlth.Sci. Ctr., New Orleans, LA, [22]Univ. of California Los Angeles Med. Ctr., Los Angeles, CA, [23]Univ. of Maryland Sch. of Med., Baltimore, MD, [24]Med. Univ. of South Carolina, Charleston, SC, [25]Univ. of Mississippi Med. Ctr., Jackson, MS, [26]Tufts Med. Ctr. and Tufts Univ. Sch. of Med., Boston, MA

Abstract:

A pervasive issue for all of precision medicine is assessing disease risk in the first-degree relatives of a proband diagnosed with a heritable disease. This issue arises because probands are frequently ill-equipped to disseminate genetic risk information to their relatives, and motivation for screening among these relatives is variable. Clinical screening of first-degree relatives is especially important for risk management in conditions like dilated cardiomyopathy (DCM), which is asymptomatic until late-phase disease. Because first presenting with late-phase disease may compromise responses to therapy, clinical cardiovascular screening to identify asymptomatic relatives with prevalent or incident DCM is critical. Early identification is especially important in Black families, in whom familial DCM is more frequent. To address this issue, the DCM Precision Medicine Study developed *Family Heart Talk*, a booklet designed to help DCM probands communicate genetic risk, and assessed its effectiveness in a multicenter, open-label, cluster-randomized, controlled trial. Between June 2016 and March 2020, 1241 probands were randomized (1:1) to receive *Family Heart Talk* (n=621) or not (n=620) within strata defined by site and self-identified race and ethnicity (non-Hispanic Black, non-Hispanic White, Hispanic). The primary outcome was completion of clinical cardiovascular screening initiated within 12 months after proband enrollment, which was prior to return of any genetic results to the family, among living first-degree relatives. Final analyses included 553 families (n=2243 first-degree relatives) in the *Family Heart Talk* arm and 566 (n=2446) in the control arm. A higher percentage of first-degree relatives completed screening in the *Family Heart Talk* arm compared to the control arm (19.6% vs. 15.9%; adjusted OR=1.32; one-sided 95% CI: 1.09 - ∞; one-sided p=0.007), and the intervention effect did not differ by race-ethnicity group (P=0.90). We conclude that *Family Heart Talk*, a booklet that can be provided to DCM patients by clinicians with minimal additional time investment, was effective in increasing clinical cardiovascular screening among their first-degree relatives.

S44. Voices from the community

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 10:45 am - 12:15 pm

ProgNbr 287. Creation of a Justice, Equity, Diversity and Inclusion Action Plan to inform the National Society of Genetic Counselors' organizational goals

Authors:

H. Zierhut[1], T. Jasper[2], J. Vento[3], M. Carey[2], D. Babu[4], A. Davis[5]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Natl. Society of Genetic Counselors, Chicago, IL, [3]Univ. of Pittsburgh, Pittsburgh, PA, [4]Integrity Content Consulting, St. Albert, AB, Canada, [5]The Exeter Group, Chicago, IL

Abstract:

Background: The National Society of Genetic Counselors (NSGC) has a goal to develop, implement and operationalize a sustainable organizational structure and culture that supports justice, equity, diversity and inclusion (J.E.D.I). NSGC engaged The Exeter Group (Exeter) in May 2020 to complete a Diversity, Equity, and Inclusion (DEI) organizational assessment. A key recommendation was the establishment of a J.E.D.I. Action Plan, a tactical plan that starts within the organization and has impact extending into the healthcare system and communities served. **Methods:**A task force of 13 members and nonmembers was convened using an application process guided by Exeter. A data-driven process to prioritize recommendations was initiated and led by Exeter. First, the Task Force rated 50 potential recommendations extracted from the DEI report and rated them on a 10-point scale to determine potential impact to NSGC and difficulty to implement. The Task Force then categorized the recommendations and discussed data, resources and metrics for each. Additionally, they identified in what year over a three year period that each recommendation should be initiated with a goal of 10-15 recommendations each for Year 1 (Do Now) and 2 (Do Next). A small group of NSGC Board members and staff convened to review the draft report and list of prioritized actions. Feedback to the Task Force was given and a second draft was reviewed and approved by the Board. **Results:**The Task Force created a three-year action plan with 38 actions that is designed to address J.E.D.I. from perspectives across staff, members, lapsed members, and the genetic counseling field. After two rounds of review and prioritization, most actions fell into Do Now (n=23) and Do Next (n=11) with only a few in Do Later (n=4). The plan is categorized around six segments: Communication and Transparency (n=6), Conference (n=6), Education and Training (n=6), Policy Changes (n=9), Partnerships and Outreach (n=6), and Wide-sweeping (n=5). **Discussion:**The Action Plan provides a public means of communicating J.E.D.I. related goals, metrics, and collectively provides NSGC with opportunities to introduce and bolster J.E.D.I. in the organization. Members, staff, and leaders will be better informed about J.E.D.I. and how their behaviors help or hinder progress. Further, inequities and disparities will be addressed through intentional policy, process, and programming changes. To the extent possible, NSGC will engage with external partners to increase J.E.D.I. in the field. With this plan, NSGC can boldly make changes to introduce and improve J.E.D.I. throughout the organization.

S45. All of Us for All of You

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 290. Genomic data in the *All of Us* Research Program: Advancing precision medicine for all

Authors:

A. Bick[1], K. Mayo[2], A. Able[2], G. Metcalf[3], L. Lichtenstein[4], S. Rura[5], R. Carroll[1], A. Musick[6], All of Us Genome Centers, All of Us Data and Research Center, All of Us Research Program; [1]Vanderbilt Univ. Sch. of Med., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Baylor Coll. of Med., Houston, TX, [4]Broad Inst., Cambridge, MA, [5]Verily Life Sci., Cambridge, MA, [6]NIH, Bethesda, MD

Abstract:

Background: The NIH's All of Us Research Program (AOURP) seeks to collect health data from a million or more participants to create a diverse research resource that accelerates precision medicine. Here we report the first release of whole genome sequencing (WGS) data on 98,622 All of Us participants. To demonstrate the quality of the data and enable its wider use, we performed a series of data harmonization and quality control procedures. We also conducted several analyses characterizing the properties of the dataset, including relatedness and genetic ancestry, as well as validating the data by replicating well established genotype-phenotype relationships.
Methods: AoURP Genome Centers generated ~30x WGS data with a harmonized CAP/CLIA validated pipeline in accordance with the program's FDA Investigational Device Exemption protocol. Genomic variants were jointly called across all samples using GATK Best Practices, except that the variants are stored in a bigquery based genomic variant store. The resulting genotype and phenotype dataset was validated using a genomewide association study for LDL and a Phenome wide genotype/phenotype replication study.
Results: After quality control procedures, the joint called WGS dataset contained ~593M unique variants across the 98,622 individuals. 80% of these participants are from communities historically underrepresented in biomedical research and ~50% are individuals from racial or ethnic minority groups. We performed a single variant genome wide association study for Low Density Lipoprotein (LDL) cholesterol and identified 9 well established loci at genome wide significance, with minimal genomic inflation. The effect estimates of these loci in AoURP were highly concordant (R2 = 0.87) with a recent LDL GWAS in the NHLBI TOPMed study (N=66,329, 56% non-European ancestry). We then looked across 3,724 variants associated with 117 electronic health record defined phenotypes from the Phenotype / Genotype Reference Map and found high overall replication rates across both in AoURP participants of European ancestry (68%) and African ancestry (46%).
Outlook: Less than five years since enrolling the first participant in May 2017, the AoURP delivered its first major release of high-quality genomic data to over 1,000 approved researchers in the AoU Researcher Workbench cloud platform and to the public through the All of Us Public Data Browser (https://databrowser.researchallofus.org/). We anticipate ongoing genomic data releases twice a year.

S45. All of Us for All of You

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 291. Considerations for meaningful collaboration with Tribal Nations and self-identified American Indians and Alaska Natives in the *All of Us* Research Program

Authors:

M. Hahn, M. Samant, R. Peterson, S. Devaney, N. Aspaas, K. Rush, K. Onyeneho, K. Watson, J. Denny, All of Us Research Program; NIH, Bethesda, MD

Abstract:

Since its national launch in 2018, the *All of Us* Research Program ("program") has consented nearly 500K diverse participants, most of whom have submitted their biospecimens, health surveys, electronic health records, and/or wearable data. These data will be used to return results to participants and for analysis by authorized researchers of the program's Researcher Workbench, a cloud-based platform enabling data analysis and collaboration. The data available for research includes whole genome sequences from nearly 100K participants, representative of ~50% diverse populations by race/ethnicity and 80% historically underrepresented in biomedical research. However, as the program consulted with Tribal leaders, data from self-identified AI/AN individuals has been withheld from researcher datasets. *All of Us* conducted formal government-to-government consultations in 2019, engaging with Tribal leaders about inclusion of AI/AN communities in population cohort studies like *All of Us*. These robust Tribal consultations led to the development of immediate, short-/long-range plans and a report centering on: Tribal sovereignty and research participation, cultural sensitivity and awareness; data access, use, and protection; governance involvement; and protecting data and preventing re-identification. The consultations resulted in 12 commitments between the program and Tribal Nations, which this presentation will highlight. All data from self-identified AI/AN individuals has been withheld until now to provide AI/AN participants time to learn more about consultations and make an informed decision about continuing in the program. For the first time ever, beginning Winter 2022, the program anticipates releasing 200K genomes and array data in the Researcher Workbench, which will include up to 15K self-identified AI/AN participants who volunteered to participate in the program. *All of Us* is implementing a detailed communication and engagement plan that will include enhanced training and guidance for authorized users of the Researcher Workbench for the purpose of promoting cultural sensitivity and emphasizing the avoidance of stigmatizing research. These efforts take into consideration impacts on AI/AN participants, program partners, and interested stakeholders and will include an assessment of responsible research with AI/AN-specific participant data. *All of Us* is committed to engaging in respectful, equitable, and ongoing dialogue and conversations with Tribal Nations and AI/AN communities.

S45. All of Us for All of You

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 292. Frequency of known pathogenic variants across ancestries in the All of Us Research Program cohort.

Authors:

E. Venner[1], K. Patterson[2], D. Kalra[1], M. Wheeler[2], Y-J. Chen[1], P. Empey[3], K. Walker[1], J. Karnes[4], S. Kalla[1], J. Smith[2], S. McGee[5], A. Radhakrishnan[2], A. Haddad[6], Q. Wang[1], G. Jarvik[7], D. Toledo[8], A. Musick[9], R. Gibbs[1]; [1]HGSC - Baylor Coll. of Med., Houston, TX, [2]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [3]Dept. of Pharmacy and Therapeutics, Univ. of Pittsburgh Sch. of Pharmacy, Pittsburgh, PA, [4]Univ. of Arizona, Tucson, AZ, [5]Univ. of Washington Genome Sci., Seattle, WA, [6]Univ. of Pittsburgh, Pittsburgh, PA, [7]Univ. of Washington Med. Ctr., Seattle, WA, [8]The Broad Inst. of MIT and Harvard, Cambridge, MA, [9]NIH All of Us Res. Program, NIH Office of the Director, Bethesda, MD

Abstract:

The All of Us Research Program (AoURP) has prioritized engaging participants from diverse backgrounds. To understand how rates of known pathogenic variants differ by ancestry group, we used an internal database of curated variants to detect known pathogenic variants and predicted genetic ancestry derived from AoURP data to group participants. Identifying ancestry-linked differences in pathogenic variant rates can reveal disparities in known pathogenic alleles and thus highlight the value of the AoURP's mode of recruitment.

The Beta release of the AoURP controlled tier data includes whole genome sequences (WGS) from 98,622 participants, of predominantly European (49%), African (23.4%), American Admixed/Latino (15.3%), East Asian (2.1%), South Asian (1.0%), Middle Eastern (0.2%) or 'Other' (9%) ancestry. An initial assessment of clinically-actionable genetic findings within the AoURP, utilizing aggregate variant data from WGS, unlinked to individual samples, grew our internal database by 20% (550 pathogenic variants). The European group had the highest rate (2.1%) of previously-known pathogenic variation, followed by Other (1.87%), African (1.48%), South Asian (1.34%), American Admixed/Latino (1.16%), Middle Eastern (1.18%) and East Asian (1.04%) respectively. Even excluding HFE, which has a very large known effect, these rates differ significantly between European, African and Admixed American/Latino groups (Chi-square: $p < 0.001$) and are largely driven by divergent pathogenic variant rates in APOB ($p = 0.0001$) and PALB2 ($p = 0.004$). Differing rates of pathogenic variation in PALB2 by ancestry have been reported previously. A comparison to rare loss of function variants found lower between-group variability, suggesting that ascertainment bias of pathogenic variants may partially explain these differences. Pathogenic variants are observed predominantly in Breast Cancer (414), Familial hypercholesterolemia (374), and Hereditary hemochromatosis (256) genes. Further comparisons to gnomAD pathogenic variant frequencies showed high-level concordance (e.g. Pearson correlation 0.99 of frequencies in the most commonly mutated European genes). Some outliers were reduced by comparing to the non-cancer or non-TopMed gnomAD subsets. Participants with ICD10 codes or survey responses indicating breast cancer history are enriched for pathogenic variants in genes related to breast cancer (32/1,653 vs 414/98,622, $p<0.00001$).

Ancestry-linked differences reveal areas of current health disparities and directs future work aimed at reducing those disparities, which will be a lasting legacy of the AoURP.

S45. All of Us for All of You

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 293. Pharmacogenomic investigation of SSRI or SNRI-induced SIADH in *All of Us*

Authors:

H. Mo[1], D. Schlueter[1], C. Zeng[1], S. Goleva[1], A. Williams[1], T. Ferrara[1], O. Stubblefield[1], J. Keaton[1], E. A. Larson[2], R. A. Wilke[2], J. Denny[1]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Univ. of South Dakota, Sioux, SD

Abstract:

Selective serotonin reuptake inhibitors (SSRI, e.g., sertraline) and Serotonin-norepinephrine reuptake inhibitor (SNRI, e.g., duloxetine) are among the most prescribed antidepressants. The syndrome of inappropriate antidiuretic hormone secretion (SIADH) can lead to clinically significant hyponatremia in patients using these medications, especially in elderly patients. Studies of SIADH during antidepressant therapy require complex phenotyping strategies built upon temporal relationship of events. The *All of Us* Research Program contains Electronic Health Record (EHR) data on 200k participants to define temporal phenotypes; nearly 100k *All of Us* participants currently also have whole genome sequencing data. We created a phenotype algorithm for drug induced SIADH defined by the presence of at least one plasma sodium measurement (Na^+) <130 mEq/L within 2 years after the mentions of an SSRI or SNRI (n = 4,096). The control group was defined by the presence of normal Na^+ level recorded within the same time frame as well as the absence of any abnormal Na^+ or hyponatremia or SIADH diagnostic codes (n = 28,631). Individual drugs showed different case-to-control ratios (p = 1.5e-17). For example, duloxetine had an adjusted odds ratio of 1.25 compared to sertraline. To assess phenotype quality, we performed a phenome-wide association study (PheWAS) with diagnostic codes recorded from 6 months prior to 2 months after each initial Na^{++} measurement. Because the PheWAS odds ratios were increased for hypovolemia, sepsis, and acute renal failure, patients with these diagnostic codes were excluded in subsequent analyses. We then performed a GWAS with our post-exclusion phenotype in a subset of individuals with overlapping WGS data (n = 756 cases, n = 9,894 controls) using HAIL. Post-exclusion data showed three significant GWAS peaks proximal to *TMEM183A* (p = 4.96e-8), *ANKRD30A* (p = 5.7e-8), and *SGCG* (p = 4.76e-8). Our study shows the potential of *All of Us* as a tool for pharmacogenomic investigation.

S46. Genetics of substance use disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 296. Cross-ancestry meta-analysis of tobacco use disorders based on electronic health record data uncovers novel loci and reveals associations with numerous health outcomes

Authors:

M. Jennings[1], S. Toikumo[2], H. Lee[3], T. Mallard[3], S. Bianchi[1], B. Pham[1], N. Courchesne-Krak[1], M. Niarchou[4], Million Veteran Program, psycheMERGE Substance Use Disorder Workgroup, D. Hancock[5], K. Xu[6], J. Smoller[3], L. Davis[4], A. Justice[7], H. Kranzler[8], R. Kember[2], S. Sanchez-Roige[1]; [1]Univ. of California, San Diego, La Jolla, CA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]RTI Interntional, Durham, NC, [6]Yale Sch. Med., New Haven, CT, [7]Yale Univ. Sch. of Med. and VA Connecticut, New Haven, CT, [8]Univ. of Pennsylvania Perelman Sch. of Med. and Crescenz VAMC, Philadelphia, PA

Abstract:

Tobacco use disorders (TUD) are the most prevalent substance use disorder in the US, with a high proportion of smokers meeting criteria for dependence. These individuals often have difficulty quitting, experience withdrawal symptoms when they stop, and continue smoking despite negative mental, social, and medical consequences. Genetic factors influence smoking behaviors and strides have been made in understanding aspects of tobacco initiation and use via genome-wide association studies (GWAS). However, due to limited sample sizes, GWAS of TUD still account for only a small amount of the variance in these traits. In addition, nicotine-related GWAS have primarily focused on individuals of European ancestry. Here we leverage access to multiple biobanks (Vanderbilt University Medical Center, Mass General Brigham, Million Veterans Program, UK Biobank) to perform a cross-ancestral meta-analysis of TUD (derived via electronic health records, EHR) in 740,361 individuals of European, African American, and Latin American ancestries. The TUD-EHR phenotypes were genetically correlated (rg=0.51-1.24) across all sites. The cross-ancestral meta-analysis identified 31 independent loci, including the nicotinic acetylcholine receptor CHRNA3/A4 gene cluster, which has been consistently associated with smoking behaviors. Other promising candidate genes include PDE4B, previously associated with smoking initiation and problematic alcohol use, and PTPRF, recently implicated in opioid addiction and impulsive behaviors. TUD-EHR was also genetically correlated with traits derived from traditionally ascertained cohorts, including nicotine dependence and smoking cessation. Surprisingly, the genetic correlation between cigarettes per day and TUD-EHR, although significant and positive, was moderate in magnitude (rg<0.5), suggesting that the genetic architecture of consumption and misuse is distinct. Lastly, we evaluated the use of TUD-EHR polygenic scores as genomic predictors of 1,335 psychiatric and medical traits, and revealed hundreds of associations, including HIV infection, heart disease, and pain. This work furthers our biological understanding of TUD and its shared genetic risk with other mental and physical traits, and establishes that EHR are useful sources of phenotypic information for studying the genetics of TUD.

S46. Genetics of substance use disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 297. Exome-wide association study in ~750,000 individuals identifies CHRNB2 as a potential drug target for smoking

Authors:

V. Rajagopal[1], D. Sharma[1], M. Kessler[1], K. Watanabe[1], J. Mbatchou[1], N. Parikshak[1], M. Kapoor[1], J. Emberson[2], R. COLLINS[2], J. Torres[2], P. Kuri[3], R. Tapia-Conyer[4], J. Alegre[3], J. Berumen[3], R. Genetics Center[1], G-R. DiscovEHR Collaboration[1], J. Marchini[1], A. Schuldiner[1], E. Jorgenson[1], M. Cantor[1], A. Baras[1], G. Coppola[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Natl. Autonomous Univ. of Mexico, Mexico City, Mexico, [4]Fundación Carlos Slim, Mexico City, Mexico

Abstract:

Background: Human genetic studies of smoking behavior have so far been limited largely to common variants. Studying rare coding variants has the potential to identify new drug targets. Here we present an analysis of associations of rare coding variants with a broad range of smoking behaviors in 120,000 to 750,000 exomes.
Methods: We analyzed six primary phenotypes—nicotine addiction (ICD10-F17), ever-smoker, heavy-smoker (>10 cig/day), former-smoker, number of cigarettes smoked per day (cig-per-day) and age started smoking, and 32 secondary phenotypes related to either smoking or smoking-related diseases. Study populations include Europeans, Admixed Americans, Africans, South-Asians and East-Asians. Genetic analyses were performed using REGENIE. We focused on functional coding variants and two types of burden masks—M1 (predicted loss of function [pLoF] variants) and M3 (pLoF + deleterious missense variants)—at five allele frequencies (singletons, 0.0001, 0.001, 0.01, 0.05). Associations with P<5e-8 were evaluated further using secondary analyses.
Results: We highlight here three major findings. 1) M3 burden in *CHRNB2* showed protective associations with heavy smoking (M3.01: OR=0.65; CI=0.56-0.76; P=1.9e-8) and multiple other phenotypes. Notably, *CHRNB2* is a known drug target of varenicline used for smoking cessation. Secondary analyses revealed a Finnish enriched (2.6 times) deleterious missense variant (p.Arg460Gly) which showed a consistent protective association with substance abuse (OR=0.72; P=0.04) in Finngen. Also, a common variant (rs2072659) near *CHRNB2* independently showed protective association. 2) M1 and M3 burdens in *CHRNA5* and *CHRNA3* showed risk associations with cig-per-day (M3.5: b=0.11; P=1.1e-16; M1.1: b=0.2; P=0.001) and multiple other phenotypes, independent of the known strong common variant signal. 3) M1 and M3 burdens in *ASXL1* and *DNMT3A* showed strong risk associations with nearly all the smoking-related phenotypes.
Discussion: We observed a striking convergence across multiple lines of evidence (common variants, rare variants, published mice studies, and known drug target) suggesting that loss of *CHRNB2* leads to reduced smoking. Our results also show that loss of *CHRNA5* and *CHRNA3* independently leads to increased smoking, further strengthening what has been previously known based on common risk variants and mice studies. *ASXL1* and *DNMT3A* are known clonal hematopoiesis of indeterminate potential (CHIP) genes and hence, the rare variants associated with smoking are likely somatic mutations suggesting that smoking accelerates the clonal expansion of CHIP mutations

S46. Genetics of substance use disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 298. Multi-omics insights into the biological mechanisms underlying gene-by-smoking and gene-by-alcohol consumption interactions

Authors:

A. Bentley[1], T. D. Majarian[2], V. Laville[3], M. R. Brown[4], D. I. Chasman[5], C. N. Rotimi[1], H. Aschard[3], C. Gu[6], A. Manning[7], CHARGE Gene-Lifestyle Interactions Working Group; [1]NIH, Bethesda, MD, [2]Broad Inst., Cambridge, MA, [3]Inst. Pasteur, Paris, France, [4]UTHlth.Houston, Houston, TX, [5]Brigham & Women's Hosp., Boston, MA, [6]Washington Univ. Sch. of Med., St. Louis, MO, [7]Massachusetts Gen. Hosp., Boston, MA

Abstract:

Though both genetic and lifestyle factors are known to influence cardiometabolic outcomes, less attention has been given to whether lifestyle exposures can modify the association between a genetic variant and these outcomes. The CHARGE Consortium's Gene-Lifestyle Interactions Working Group has recently published investigations of genome-wide gene-environment interactions in large multi-ancestry meta-analyses with a focus on cigarette smoking and alcohol consumption as lifestyle factors and blood pressure and serum lipids as outcomes. Further description of the biological mechanisms underlying these statistical interactions would represent a significant advance in our understanding, yet accessing and harmonizing individual-level genetic and 'omics data is challenging. Here we demonstrate the coordinated use of summary-level data for gene-lifestyle interaction associations on up to 600,000 individuals, differential methylation data, and gene expression data for the characterization and prioritization of loci for future follow-up analyses. Using this approach, we identify 48 genes for which there are multiple sources of functional support for the identified gene-lifestyle interaction. We also identified five genes for which differential expression was observed by the same lifestyle factor for which a gene-lifestyle interaction was found. For instance, in gene-lifestyle interaction analyses, the T allele of rs6490056 (*ALDH2*) was associated with higher systolic blood pressure, and a larger effect was observed in smokers compared to non-smokers. This allele is associated with decreased expression of *ALDH2*, which is part of a major oxidative pathway. Increased expression of *ALDH2* is observed among smokers, potentially as a response to the oxidants in cigarette smoke. Oxidative stress is known to contribute to worsening blood pressure. Together these data support the hypothesis that rs6490056 reduces expression of *ALDH2*, which raises oxidative stress, leading to an increase in blood pressure, with an even stronger effect among smokers, in whom the burden of oxidative stress is greater. Other genes for which the aggregation of data types suggest a potential mechanism underlying gene-lifestyle interactions include: rs3761743 (*GCNT4*) × current smoking on HDL levels, rs77810251 (*PTPRZ1*) × ever-smoking and HDL levels, rs4135300 (*SYN2*) × current smoking on pulse pressure, and rs10849962 (*TMEM116*) × ever-smoking on mean arterial pressure. In conclusion, this work demonstrates the utility of careful curation of summary-level data from a variety of sources to prioritize gene-lifestyle interaction loci for follow-up analyses.

S46. Genetics of substance use disorders

Location: Conv Ctr/Room 502/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 299. Polygenic scores for tobacco use provide insights into disease associations in a diverse EHR-linked biobank

Authors:

V. Venkateswaran[1], A. Bhattacharya[1], K. Boulier[1], Y. Ding[1], R. Johnson[1], B. Pasaniuc[2]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abstract:

Tobacco use contributes to significant mortality and morbidity, causing over 7 million deaths annually. Genome-wide association studies (GWAS) have identified several single nucleotide polymorphisms (SNPs) associated with traits such as tobacco smoking behaviors and nicotine dependence. Given the polygenic nature of this phenotype, a polygenic score (PGS) that sums up the effects of multiple variants will capture a larger proportion of the genetic predisposition to tobacco use than individual variants alone. Here, we use a PGS for tobacco use disorder derived from UK Biobank data to examine the pleiotropic effects of variants associated with tobacco use within the UCLA ATLAS biobank, a diverse EHR-linked biobank with extensive de-identified phenotypic and demographic information (N>60,000). First, we find that tobacco use-PGS is correlated with smoking history (ever/never smoker binary variable) across different genetically inferred ancestries (GIAs) including European American, Hispanic Latino American, and East Asian American ancestries (P-values: <0.0001, 0.0062, 0.0071). Next, we investigate tobacco use-PGS correlation with 1,847 EHR-derived traits across a wide spectrum of diseases/phenotypes (i.e., phecodes) using a phenome-wide association (PGS-PheWAS) analysis, meta-analyzed across 4 GIAs. We find 21 significant associations at Bonferroni-adjusted $P < 0.05$ between traits and tobacco use-PGS after adjusting for age, sex, first 5 principal components, and health insurance. The top phecodes associated with the PGS were obesity, chronic bronchitis, substance addiction, coronary atherosclerosis, and congestive heart failure. Next, we focused on GIA-specific analyses: PGS showed additional associations with type 2 diabetes, alcoholism, and senile dementia in European American GIA. The PGS demonstrates significant associations with obesity, alcoholism, senile dementia, and hypertension in ‘never-smokers’, suggesting that tobacco use disorder may share genetic architecture with obesity and alcoholism.
Overall our study provides an extensive investigation of tobacco use-PGS across a wide range of traits derived from EHR; PGS recapitulates many known phenotypic associations of tobacco use suggesting that the genetic predisposition to tobacco use partly influences disease associations and comorbidities. Additionally, we detect new associations with tobacco use-PGS across the phenome, suggesting new risk factors or sequelae for tobacco use.

S47. Novel methods for rare variants

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 302. Heritability and genetic architecture of rare coding variation across 394,000 exomes

Authors:

A. Nadig[1], D. Weiner[1], K. Jagadeesh[1], K. Dey[1], B. Neale[2], E. Robinson[2], K. Karczewski[1], L. O'Connor[1]; [1]Broad Inst., Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA

Abstract:

Common diseases and traits are influenced by thousands of common genetic variants, and recently, large-scale sequencing studies have identified hundreds of rare-variant associations as well. Rare-variant associations can have much larger effect sizes, but it is unclear how many of them there are, how much heritability they explain, and more generally how their genetic architecture compares with common associations. We developed Burden Heritability Regression (BHR) to quantify the heritability explained by rare and ultra-rare coding variants. We applied this method to exome-sequencing data from 24 common diseases and traits, and systematically compared the heritability and genetic architecture of rare and common variation.
BHR estimates burden heritability, which is the phenotypic variance explained by burden of rare (MAF<1e-3) alleles. It regresses gene-wise burden association statistics on “burden scores,” distinguishing genuine heritability from confounding (like LDSC; Bulik-Sullivan 2015 Nat Genet). We confirmed that it produces unbiased estimates in simulations.
We first applied BHR to exome sequencing data for 22 UK Biobank traits (N=394k; Karczewski 2022 medRxiv). We estimate that rare and ultra-rare loss-of-function variants explain 1.0% (SE = 0.1%) of phenotypic variance on average, much less than common variants. Common and rare variants implicate the same cell types, with similar enrichments; they have pleiotropic effects on the same trait pairs, with similar genetic correlations; and they partially colocalize at individual genes. However, rare-variant burden heritability is concentrated within far fewer genes (median: 7 significant genes explain 18% of burden heritability), indicating that rare-variant architecture is much less polygenic.
Applying BHR to exome-sequencing data for schizophrenia and bipolar disorder (Singh 2022 Nature, Palmer 2022 Nat Genet), we find that their burden heritability is unusually high (3.6% and 5.1%, respectively). GnomAD constrained genes are enriched for burden heritability across all traits (median:3.7x), but the enrichment is especially strong for schizophrenia and bipolar disorder (16.9x and 6.5x, respectively).
Our results show that there are a tractable number of trait-associated genes to discover by studying rare variants, that the biological insights derived from these genes should be applicable to common variant-driven forms of disease, and that rare coding variants are unlikely to explain substantial missing heritability or improve population risk stratification. More broadly, our results unify rare- and common-variant associations with common diseases and complex traits.

S47. Novel methods for rare variants

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 303. Rare variant polygenic risk scores

Authors:

J. McRae, P. Fiziev, K. Farh; Illumina, Inc, Foster City, CA

Abstract:

Although individual GWAS variants confer effects that tend to be too mild for clinical actionability, polygenic risk scores (PRS) combining signal from hundreds to millions of common variants have demonstrated significant efficacy for predicting patients at phenotypic extremes who are at high risk for disease. However, existing common variant PRS models have largely excluded rare variants due to challenges in interpreting variants of unknown significance and imprecision in their effect size estimation. Here we propose a complementary, rare variant PRS model, based on a weighted sum of rare deleterious variants from multiple phenotype-associated genes.

We constructed rare PRS models using UK Biobank exome data, and compared with common PRS models from UK Biobank GWAS data. While rare PRS models explained a smaller fraction of the phenotypic variance than common PRS models, the rare PRS models are particularly good at identifying individuals who were outliers for each trait, that is, the individuals most at risk of developing disease, and who might benefit most from clinical intervention. We found individuals at the extremes of the rare PRS were much better at discriminating these phenotypic outliers than individuals at the extremes of the common PRS. Furthermore, the rare PRS identified more individuals with >4-fold disease risk compared to the common PRS.

Our rare PRS models are portable between cohorts and ethnicities. We validated our rare PRS models from UK Biobank in an independent exome-sequenced cohort. The rare PRS models were also robust to differences in ancestry, as the same effects were observed in individuals with European ancestry vs individuals with non-European ancestry.

S47. Novel methods for rare variants

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 304. PheWAS of ultra-rare deleterious variant burden in the UK Biobank: Insights into hundreds of complex traits with high locus heterogeneity

Authors:

K. Carss[1], O. Burren[2], Q. Wang[3], E. Gardner[4], S. Petrovski[5], M. Hurles[6]; [1]AstraZeneca, Melbourn, United Kingdom, [2]Discovery Sci., R&D, AstraZeneca, Royston, United Kingdom, [3]AstraZeneca, Chapel Hill, NC, [4]Wellcome Sanger Inst., Hinxton, United Kingdom, [5]AstraZeneca, Fulbourn, United Kingdom, [6]Wellcome Sanger Inst., Cambridge, United Kingdom

Abstract:

The selection coefficient (s_{het}) burden score reflects an individual's exome-wide burden of deleterious variants, and is negatively correlated with fecundity.[1] Here we tested the hypothesis that a modified version of s_{het} burden score could be associated with a wider range of human traits. We calculated s_{het} burden score for each of 393,076 UK Biobank participants using private (minor allele count = 1) protein-truncating variants (PTVs) and large exon-overlapping deletions. We conducted a phenome-wide regression-based association study of s_{het} burden score against 13,540 binary endpoints and 1,418 quantitative traits, correcting for age, sex, ancestry, deprivation, and synonymous variant burden.
We identified a total of 357 binary and 58 quantitative traits significantly ($p<1x10^{-8}$) associated with s_{het} burden score. These span multiple phenotypic categories and include novel as well as previously reported rare variant associations. Examples include developmental disorders ($p=9.3x10^{-21}$; OR=65), diabetes ($p=1.4x10^{-31}$; OR=2.4), and hypertension ($p=1.8x10^{-22}$; OR=1.7). Higher s_{het} burden score is also associated with fewer children fathered, corroborating previous work ($p=1.3x10^{-19}$; beta= -0.26), as well as more health dissatisfaction ($p=5.9x10^{-44}$; beta= 0.26) and longer time to complete a pair matching game ($p=2.1x10^{-56}$; beta= 0.33), which may reflect lower fluid intelligence. s_{het} burden score also impacts overall survival ($p=1x10^{-22}$; HR= 2.05). By masking genes significantly associated with these phenotypes at the gene-level, we can also disentangle monogenic from polygenic disease associations, and find that the majority of associations are likely polygenic.
Overall, these results yield known and novel insights into the genetic architecture of hundreds of complex traits, particularly those with high locus heterogeneity and those driven by variants too rare to be assessed at the gene or variant level.
1) Reduced reproductive success is associated with selective constraint on human genes. Gardner EJ et al, Nature. 2022; 603(7903):858-863

S47. Novel methods for rare variants

Location: Conv Ctr/Room 515/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 305. Exome sequencing analysis of 14 diseases studied by the GBMI identifies likely effector genes of GWAS signals

Authors:

L. Gurski, M. Kessler, M. Suciu, J. Horowitz, A. Locke, A. Ayer, K. Siminovitch, G. Abecasis, M. Ferreira, Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration; Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

The Global Biobank Meta-Analysis Initiative (GBMI) is a multi-biobank collaboration across 23 biobanks. Recently, the GBMI released summary statistics from GWAS of 14 diseases: asthma (202 loci), glaucoma (64 loci), gout (55 loci), chronic obstructive pulmonary disease (COPD; 44 loci), venous thromboembolism (VTE; 32 loci), thyroid cancer (24 loci), aortic aneurysm (22 loci), heart failure (18 loci), stroke (13 loci), idiopathic pulmonary fibrosis (IPF; 12 loci), abdominal acute appendicitis (10 loci), appendectomy (5 loci), uterine cancer (4 loci), and cardiomyopathy (1 loci). A challenge with common variant associations from GWAS is that it is often unclear which are the likely effector genes underlying the observed associations. Associations with rare coding variants provide independent information that may help prioritize likely effector genes at GWAS loci. We performed association analyses with rare coding variants from exome sequencing across the 14 diseases studied by the GBMI, across a total of 680,000 individuals from 4 cohorts and 5 ancestries. For each GWAS locus identified by the GBMI, we selected genes based on distance (nearest gene) and co-localization with regulatory (expression and protein QTL), non-synonymous or enhancer-overlapping variants, and determined if any were associated with the same trait in our exome-sequencing analysis of rare coding variants, after accounting for the number of genes tested across all loci. We only considered associations with rare (MAF<1%) predicted loss-of-function and deleterious missense variants, tested on aggregate through gene burden tests. Across the 506 GWAS loci identified for the 14 diseases, rare variant associations that remained significant after correcting for multiple testing were observed for 17 genes in 17 (4%) loci: 6 for gout (SLC22A12, ABCG2, SLC2A9, PKD2, LRP2, SLC22A11; located in 11% [6/55] of GWAS loci for gout), 3 for glaucoma (LTBP2, ANGPTL7, LMX1B; 5%), 2 for VTE (PROC, STAB2; 6%), 2 for asthma (FLG, IL33; 1%), 1 for COPD (FGF10; 2%), 1 for thyroid cancer (CHEK2; 4%), 1 for IPF (TERT; 1%), and 1 for heart failure (SPATS2L; 3%). Of these, a risk lowering association with rare non-synonymous variants was observed for five genes: IL33 and asthma (MAF=0.006; OR=0.70); TERT and IPF (MAF=0.004 OR=0.84); ANGPTL7 and glaucoma (MAF=0.008, OR=0.77); SLC2A9 (MAF=0.001, OR=0.39) and SLC22A12 (MAF=0.003, OR=0.33) and gout. For all 17 genes, published biological evidence supports a role in the respective associated disease. Through analysis of exome sequencing data, we identified likely effector genes of GWAS signals at 17 out of 506 loci reported by the GBMI across 14 diseases.

S48. Reading between the reads: Getting the most out of sequencing results

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 308. Seeing beyond the target: leveraging off-target reads in targeted clinical tumor sequencing to identify prognostic biomarkers and survival

Authors:

Y. Patel[1], S. Mangul[1], J. J. Brito[1], S. Groha[2], N. Zaitlen[3], A. Gusev[4]; [1]Dept. of Clinical Pharmacy, Sch. of Pharmacy, Univ. of Southern California, Los Angeles, CA, [2]Dana Farber Cancer Inst. and Harvard Med. Sch., Boston, MA, [3]Dept.s of Neurology and Computational Med., Univ. of California Los Angeles, Los Angeles, CA, [4]Dana-Farber Cancer Inst., Boston, MA

Abstract:

Tumor sequencing is becoming a standard component of clinical care, providing essential information for selecting amongst treatment options and providing prognostic value. To reduce costs, tumor sequencing platforms sequence exons from a small number of known cancer genes. However, capture technologies are imperfect, and off-target reads are produced across various technologies. A recent Nature publication by Bentham et al. estimated T cell fractions inferred from whole-exome sequenced tumor samples; we extend this work with integrated genomics methods to develop a robust and scalable software platform (SBT: Seeing Beyond the Target) that mines discarded components of clinical sequences to produce estimates of a rich set of omics features including rDNA and mtDNA copy number, microbial species abundance, and T and B cell receptor sequences. We show that SBT can uncover components of the tumor microenvironment that may serve as prognostic biomarkers. We performed benchmarking using whole exome and transcriptome data to show SBT can accurately estimate T and B cell receptor sequence diversity; microbial, ribosomal, and mitochondrial profiles. We validated the accuracy of SBT via an analysis of a tumor panel cohort of 2,920 lung adenocarcinomas with sequencing from the Dana-Farber *Profile* cohort - a prospective collection of patient biopsies - replicating 6 published discoveries that required specifically designed experiments. We replicated known associations of somatic events in *TP53* with changes in rDNA ($p=0.012$); as well as diversity of BCR and TCR repertoires with the biopsy site ($p=2.5x10-6$, $p<10-20$). We observed striking differences in EGFR mutant lung cancers versus wild-type, including higher rDNA copy number and lower immune repertoire diversity. Integrating clinical outcomes, we identified significant prognostic associations with overall survival, including SBT estimates of 5S rDNA ($p=1.9x10-4$, hazard ratio = 1.22) and TCR diversity ($p=2.7x10-3$, hazard ratio=1.77). Both novel survival associations replicated in 1,302 breast carcinoma and 1,651 colorectal cancer tumors. We anticipate that feature estimates derived by SBT will yield novel biomarker hypotheses and open research opportunities in existing and emerging clinical tumor sequencing cohorts.

S48. Reading between the reads: Getting the most out of sequencing results

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 309. Exome versus panel germline sequencing for pediatric cancer predisposition syndromes: Is more always better?

Authors:

L. Desrosiers[1,2], T. Wang[1], J. Reuther[3], G. Miles[1,2], H. Dai[1,4], E. Jo[1], R. Raesz-Martinez[1,2], A. Thomas[5], E. Berenson[6], J. Corredor[7], K. Nugent[1], R. Wyatt Castillo[6], R. Althaus[5], R. Littlejohn[1], S. Scollon[1,2], G. Tomlinson[6], J. Gill[7], J. Bernini[1,8], K. Vallance[5], T. Griffin[1,9], C. Eng[1,4], S. Kulkarni[1,4], S. Hilsenbeck[1], A. Roy[1,2], A. McGuire[1], D. Parsons[1,2], S. E. Plon[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Invitae, San Francisco, CA, [4]Baylor Genetics, Houston, TX, [5]Cook Children's Hosp., Fort Worth, TX, [6]UT Hlth.San Antonio, San Antonio, TX, [7]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [8]Vannie Cook Children's Clinic, McAllen, TX, [9]Children's Hosp. of San Antonio, San Antonio, TX

Abstract:

Background: Clinical genomic testing ranging from single gene or panel analysis to genome sequencing are increasingly being used in the care of childhood cancer patients. Data are needed regarding the relative clinical utility of available tests. **Methods:** The Texas KidsCanSeq study enrolled pediatric cancer patients from 2018 to 2021 at six institutions as part of the NIH Clinical Sequencing Evidence-generating Research consortium. Clinical germline testing for subjects included a clinical exome and pediatric cancer panel test; only the latter reported copy number variation (CNV) and larger indels. Results were categorized by the presence of pathogenic/likely pathogenic (P/LP) variants in dominant and recessive cancer predisposition genes and those dominant “pediatric actionable” genes with cancer surveillance recommended in childhood. McNemar’s test was performed to compare the results of exome and panel. **Results:** Germline exome and panel testing was completed for 581 of 626 enrolled participants. Cancer P/LP variants in 49 genes were reported in 103/581 (17.7%) cases in at least one platform with no significant differences in frequency observed for race or Hispanic ethnicity. Exome and panel results with cancer P/LP variants were concordant in 44 cases including 3 mosaic diagnoses and discordant in 59 cases. The majority of discordant cases (52/59, 88.1%) were variants reported only by exome in genes not on the panel. The difference in diagnostic reports between exome and panel was statistically significant ($p<0.0001$). Of 30 different genes reported on exome only, *CHEK2* (n=7), *MUTYH* (n=6) and *SBDS* (n=4) were most frequent, with 18 genes reported one time. The 7 P/LP variants detected only by the panel were CNVs or indels in genes covered on both tests. Of the 103 cases with cancer P/LP variants 53 (51.5%) involved pediatric actionable genes, of which, 16 were discordant between platforms. All 7 panel-only variants in the study were pediatric actionable (e.g. *TP53* deletion) but only 9 of 52 exome-only variants were pediatric actionable (e.g. *CDKN1B* nonsense). The differences between platforms for actionable variants was not statistically significant ($p=0.6171$). **Conclusion:** Fewer than half of reported P/LP cancer cases were reported on both exome and panel. Most discordant results were in a large number of genes not covered by the pediatric cancer panel initially designed for tumor analysis, but a minority of actionable variants were CNVs and indels detected by panel-only. Careful optimization of gene content and detectable alteration types is critical for maximizing diagnostic yield and clinical utility for genetic analysis of pediatric cancer patients.

S48. Reading between the reads: Getting the most out of sequencing results

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 310. Integrated DNA and RNA analysis of hereditary cancer associated genes: More answers for more patients

Authors:

E. Esplin, B. Heald, S. Nielsen, D. Pineda-Alvarez, V. Carlton, J. J. Vincent, K. Krempely, H. Kang, N. Kamps-Hughes, L. Mette, S. Michalski, **K. Nykamp**, R. Nussbaum, L. Fresard; Invitae, San Francisco, CA

Abstract:

Background Identification of pathogenic/likely pathogenic (P/LP) germline variants is important to clarify a patient's cancer risk, inform management, guide treatment choices, and indicate cascade testing. ~14% of P/LP variants cause disease through misplicing, with >90% of these disrupting the canonical donor and acceptor splice sites. Alternatively, 69% of variants predicted to affect splicing by commonly used algorithms are located outside of the canonical sites and, without functional and clinical evidence, can only be classified as variants of uncertain significance (VUSs). RNA sequencing can help classify variants predicted to alter RNA splicing, as well as uncover splice-altering variants in regions outside of the standard reportable range of targeted DNA panels (all coding exons ±20bp of flanking intronic sequence). Here we report the resolution of VUSs that potentially impact splicing and the discovery of intronic variants using RNA sequencing performed at a single commercial laboratory.
Methods RNA-sequencing was performed on cDNA from leukocyte mRNA for 63 transcripts from an 84 gene multi-cancer panel. RNA data were used as functional evidence in our classification algorithm Sherloc to reclassify potential splicing VUSs to benign/likely benign (B/LB) or P/LP. Aberrant splicing was also used to discover variants outside of the reportable range. Data were stratified per patients' self-reported ancestries.
Results 19,327 unique patients underwent RNA analysis; of these, 3,784 (19.6%) had DNA variants predicted to affect splicing. After including positive and negative RNA results as functional evidence for interpretation, 1,227 (6.3%) of 19,327 patients had a VUS downgraded to B/LB and 78 (0.4%) patients had a VUS upgraded to P/LP. There were significantly higher rates of VUS reclassification for individuals among non-White populations, with self-reported Hispanic (11.2%, p=1.34e-18) and Black/African American (12.9%, p=1.95e-28) ancestries compared to Non-Hispanic White populations (4.8%). 42 (0.2%) patients had intronic variants outside our reportable range: 32 were classified as VUSs, 3 LP, 5 P, and 2 B.
Conclusions RNA analysis helped reclassify potential splice-altering variants in hereditary cancer syndrome genes in 6.7% of patients undergoing testing, with the majority being downgraded to B/LB. This approach was particularly effective for variants in patients from traditionally underrepresented populations, for whom VUSs have been shown to be more frequent than in Non-Hispanic White populations and a potential source of healthcare disparity. The rate of variant discovery outside our reportable range was very low.

S48. Reading between the reads: Getting the most out of sequencing results

Location: Conv Ctr/West Hall A/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 311. Whole genome sequencing in rare disease diagnosis: Update from the completion of the UK 100,000 Genomes Project

Authors:

A. Rendon[1], D. Kasperaviciute[1], S. Walker[1], K. Savage[1], K. Smith[2], A. Rueda[3], D. Smedley[4], M. Brown[1], R. Mein[5], A. Pickard[5], M. Caulfield[4], E. Thomas[1], Z. Deans[6], S. Hill[5], R. Scott[1], Genomics England Research Consortium; [1]Genomics England, London, United Kingdom, [2]AstraZeneca, Melbourn, United Kingdom, [3]Exact Sci., Oxford, United Kingdom, [4]Queen Mary Univ. of London, London, United Kingdom, [5]NHS England and Improvement, London, United Kingdom, [6]Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract:

The 100,000 Genomes Project explored the role of whole genome sequencing (WGS) in routine healthcare in the UK's National Health Service (NHS) for patients with rare disease or cancer, and has provided actionable finding to participants. To empower research, the genomic and health data are available via a secure research environment.
Over 73,000 participants from 36,000 families with more than 300 different rare diseases were recruited across the UK. WGS and automated variant prioritisation were performed by Genomics England. Analyses were returned to NHS clinical laboratories for evaluation and reporting to clinicians and then their patients. Variants were triaged based on predicted consequence, gene-disease associations, allele frequency, and segregation, and complemented with Exomiser scores. Prioritised variants were interpreted clinically by NHS teams, with access to the entire genome available for further exploration if required. We developed knowledge bases for gene-disease associations (PanelApp; >6,000 gene associations) and for prioritised and interpreted variants (Clinical Variant Ark; >3.5M variants) to empower these processes.
Overall diagnostic yield is currently 20%, varying according to condition. Example yields for some conditions with >1,000 cases include: cystic kidney disease 56%, intellectual disability 27%, and ultra rare monogenic disorders 14%. Small variants accounted for 98% of diagnoses, with the rest comprised by CNVs, short tandem repeat expansions, complex structural variants and uniparental disomies. Review of 900 variants missed by the triaging algorithm revealed that 66% of missed cases resulted from new gene disease associations or unexpected modes of inheritance. The remaining 34% were missed due to consequence type, variant quality, or allele frequency cut-off in that order. Exomiser recovered 37% of these variants (top 3; score >0.75).
To enhance yield and understand sources of new diagnoses, we, together with the research community, have undertaken cohort-wide genotype-driven approaches. To date, over 1,000 new putative diagnoses have been identified and returned to NHS labs, demonstrating that combining clinical and research activities lead to further participant benefit.
Building from the successes of the project, WGS has been commissioned as part of routine clinical care in the NHS Genomic Medicine Service in England. Since Nov 2020, over 8,000 cases (>18,000 genomes) with rare diseases have been analysed. The service has benefited extensively from the learning from the 100,000 Genomes Project through updates to pipelines and scientific knowledge.

S49. RNA based mechanisms in neuropsychiatric disorders

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 314. Isoform-level transcriptome-wide association studies uncover novel mechanisms underlying genetic associations with neuropsychiatric disorders

Authors:

A. Bhattacharya, M. Kim, C. Wen, C. Jops, D. D. Vo, J. Hervoso, B. Pasaniuc, M. J. Gandal; Univ. of California, Los Angeles, Los Angeles, CA

Abstract:

Integration of gene expression with GWAS has prioritized underlying transcriptomic mechanisms for genetic associations with disease, using colocalization and transcriptome-wide association studies (TWAS). Alternative splicing and resulting isoform diversity are extensive in the brain, under genetic control, and have been implicated in neuropsychiatric disorders. Thus, isoforms are the fundamental unit underlying the SNP to trait relationship. However, they have not been systematically studied in gene-level colocalization studies and TWAS, missing isoform-specific mechanisms when studying only genes. For example, consider a gene with multiple isoforms but only one with a trait effect; studying the total expression of this gene may miss the isoform-specific effect. Here, we present isoform-level TWAS (isoTWAS), a scalable and hierarchical framework to map trait associations to the *cis*-genetic component of isoform expression. First, isoTWAS trains multivariate predictive models of expression of a gene's isoforms, summing to total gene expression. Then, it integrates these models in a hierarchical testing and fine-mapping framework to prioritize a set of isoforms best explaining the genetic association. Using simulations and data from the Genotype-Tissue Expression Project (GTEx) and PsychENCODE Consortium (PEC), we demonstrate that by explicitly modeling isoform expression, isoTWAS greatly improves genetic prediction of total gene expression over gene-level TWAS (>25% increase in prediction R^2). Hierarchical testing also improves the power to detect trait associations. When genetic effects vary across isoforms of a gene, power increases by 40-50% over gene-level TWAS without inflated type I error. Next, we apply isoTWAS to large-scale transcriptomic reference panels from developing and adult human brain from PEC and GWAS summary statistics from 5 psychiatric disorders: schizophrenia (SCZ), bipolar disorder, Alzheimer's disease (AD), autism spectrum disorder, and attention deficit hyperactivity disorder. We identify over 200 distinct isoform-level associations (Bonferroni-adjusted $P < 0.05$ and in fine-mapped 90% credible sets). isoTWAS detects many associations that TWAS misses, including isoforms of *AKT3* and *SFMBT1* associated with SCZ and *APOC1* with AD. Lastly, in long-read RNA-seq data from adult brain ($N = 16$), we validate isoform-specific genetic effects underlying known neuropsychiatric GWAS loci using allelic imbalance analyses. This work leverages a comprehensive framework for isoform-centric interrogation of disease mechanisms to uncover previously hidden mechanisms underlying neuropsychiatric disorders.

S49. RNA based mechanisms in neuropsychiatric disorders

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 315. Spliceosome malfunction causes neurodevelopmental disorders with autistic features

Authors:

D. Li, the U2AF2 consortium, the PRFP19 consortium, the RBFOX1 consortium, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

The pre-mRNA splicing is a highly coordinated and precise process that involves numerous trans pre-mRNA binding protein factors. We present pathogenic variants in two genes, *U2AF2* and *PRPF19,* encoding core spliceosome subunits to guide the early stage of splice-site choice and activate spliceosome, respectively. By exome/genome sequencing and international matchmaking, we identified 43 unrelated individuals with 23 unique *de novo* missense variants in *U2AF2* (within or around RNA-binding domains) and four individuals with *de novo* novel *PFPF19* variants. Of 23 *U2AF2* variants, six were recurrent in 26 individuals, representing mutation hotspots of U2AF2. Detailed clinical assessment of the affected individuals with *U2AF2* or *PFPF19* variants showed consistent phenotypical presentations. *In vitro* assays demonstrated eight *U2AF2* variants affected splicing and RNA binding affinity. Crystal structures of two U2AF2 mutants were determined on X-ray, which indicated that hydrogen bonds with the polypyrimidine tract nucleotides were disrupted by the bulky residue replacement or the shortened length of the side chain. Pluripotent stem cell isogenic model was established with CRISPR knock-in for two *U2AF2* recurrent variants. Neurite length and branching number in both mutants were reduced relative to wildtype (WT), highlighting its essential role in neuronal differentiation. Multiple neural-specific RNAis of *Drosophila* orthologues *U2af50* and *Prp19* led to lethality, mushroom body (MB) patterning defects, seizures, and social deficits. Remarkably, these phenotypes seen in flies can be rescued by WT human *U2AF2* or *PRPF19*, whereas variants identified have only partial or no rescue effect, highlighting the pathogenicity of these variants. RNA-Seq of fly RNAi brains revealed three downstream effectors (including Rbfox1) which could rescue MB structural defects. Guided by these functional data, reanalysis of clinical negative exome data revealed a *de novo* variant p.(Arg118Gln) in *RBFOX1* in a patient with neurodevelopmental disorder. With the assistance of GeneMatcher, we recruited six additional patients with p.(Arg118Gln) being recurrent in three. Previous studies suggest that *RBFOX1* LoF variants may act as susceptibility factors for autism with incomplete penetrance (60%). All identified *RBFOX1* variants occurred *de novo* in the RNA-binding domain, which is highly intolerant to variants, suggesting missense variants in *RBFOX1* are associated with a neurodevelopmental disorder through modulating RNA binding affinities and specificities. This study extends the tally of causative genes in neurodevelopmental disorders to include three splicing factors.

S49. RNA based mechanisms in neuropsychiatric disorders

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 316. The dynamic FMRP-RNA interactome across the human lifespan and brain regions

Authors:

A. Lee, X. Guo, Y. Li, P. Jin; Emory Univ., Atlanta, GA

Abstract:

Fragile X syndrome (FXS) is the most common inherited cause of intellectual disability and the leading monogenic cause of autism spectrum disorder (ASD). FXS is the clinical manifestation of the loss of functional **fragile X messenger ribonucleoprotein (FMRP)**. FMRP is an RNA binding protein (RBP) enriched in the brain with its targets involved in diverse cellular processes. Crosslinking and immunoprecipitation followed by high-throughput sequencing (CLIP-seq) allows for the identification of RNA targets directly bound by RBPs. Previous CLIP studies have revealed binding targets of FMRP both shared by and unique to mice and/or humans limited to a stage of life or development. To determine whether the RNA targets of FMRP change throughout the human lifespan and differ across brain regions, **we profiled the RNAs bound by FMRP in the postmortem human brains of different ages and regions using enhanced CLIP-seq (eCLIP-seq).** Postmortem human brain tissue specimens of the dorsolateral and medial prefrontal cortex (Brodmann area 9) and caudate nucleus in three different age groups (0-year-olds, 30s, and 80s) were used for eCLIP. We identified a total of 9,348 FMRP RNA targets across samples from 9 brains in the 6 different age-by-region groups. We found that the **FMRP-RNA interactomes are dynamic in an age- and brain region- specific manner**. 186 targets were unique to the 0-year-old BA9 group, 374 to the 30s BA9, 151 to the 80s BA9, 133 to the 0s caudate, 53 to the 30s caudate, and 909 to the 80s caudate group. Distinct features of these group-specific targets were observed by gene ontology analyses. We also identified FMRP binding targets that are known to be associated with a high risk of ASD,
including *ANKRD11*, *ASH1L*, *AUTS2*, *CUL3*, *DSCAM*, *FOXP1*, *MBD5*, *MYT1L*, *NLGN4X*, *NRXN1*, *RAI1*, *SHANK2*, *TANC2*, *TL K2*. Intriguingly, the subsets of targets overlapping with ASD-associated genes display age- and brain region-dependent binding by FMRP as well. Our results also revealed individual variations in the FMRP binding RNA target profiles. We have performed whole-genome sequencing and are currently investigating the impact of genetic variation on FMRP-RNA interactions. Our study demonstrates the dynamic spatial and temporal properties of the FMRP-RNA interactome.

S49. RNA based mechanisms in neuropsychiatric disorders

Location: Conv Ctr/Petree D/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 317. Novel genes associated with the severe autism subphenotype disproportionate megalencephaly

Authors:

S. Nishizaki[1], N. Mariano[2], G. La[2], C. W. Nordahl[1], D. G. Amaral[1], M. Dennis[2]; [1]Univ. of California Davis MIND Inst., Davis, CA, [2]Univ. of California Davis, Davis, CA

Abstract:

Among autistic individuals, a subphenotype with brain enlargement disproportionate to height (autism with disproportionate megalencephaly - ASD-DM) seen at 3 years of age is associated with co-occurring intellectual disability and poorer prognoses later in life. However, little is known about the genetic factors contributing to ASD- DM. In this study we aim to identify additional ASD-DM associated genes to better define the genetic etiology of this subphenotype of autism. We identified de novo variants potentially contributing to ASD-DM observed in probands using trio and quad whole genome sequencing data from 11 families from the MIND Institute's Autism Phenome Project (APP) and 659 families from the Simons Simplex Collection (SSC) cohorts (representing 35% of SSC cases). Through this analysis we identified 119 putative loss-of-function de novo variants and one gain-of-function (duplication), overlapping a total of 118 genes, including several with previous associations with ASD-DM (CHD8, PTEN, and KMT2E). These genes are enriched for GO biological processes known to be disrupted in ASD, including cell cycle processes and histone proteins. We tested impacts of knocking out ten zebrafish gene orthologs using CRISPR gene editing and assaying morphometric features (e.g., head-size) via automated morphometric imaging. Through screening thousands of larvae, we found that two ASD-DM-candidate genes exhibited significant alterations to head-size, a proxy for brain size, in our model system. This included the gene YTHDF2, a key component of the N6-methyladenosine (m6A) modified RNA degradation complex, which we identified as partially duplicated in an ASD-DM proband. Interestingly, zebrafish knockout leads to microcephaly—matching published results of a conditional mouse knockout—while overexpression of the human YTHDF2 mRNA into zebrafish embryos leads to macrocephaly. This methodology represents a promising pipeline to identify ASD-DM candidate genes, enabling further advancement in the identification of relevant disease pathways, potential genetic therapy targets, and early detection markers.

S50. The current environment for gene-environment interactions

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 320. Using polygenic scores to detect gene-environment interactions associated with human complex traits

Authors:

L. Poyraz[1], A. Marderstein[2], A. Clark[1]; [1]Cornell Univ., Ithaca, NY, [2]Stanford Univ., Menlo Park, CA

Abstract:

In stark contrast to model organism studies, gene-environment interactions (GxE) have been a challenge to detect in human cohorts. While single-SNP GxE effects are typically small, the presence and direction of GxE may remain constant across causal variants involved in the same pathways. Hence, approaches that aggregate the GxE signal across many small-effect variants present a powerful alternative for identifying GxE. Here, we used a polygenic score (PGS)-based approach to evaluate how lifestyle factors modulate the genome-wide contribution to 7 complex traits, including body mass index, height, serum urate levels, type 2 diabetes (T2D), coronary artery disease, polycystic ovary syndrome, and gout.
First, we constructed PGS using summary statistics from previous GWAS meta-analyses and calculated optimized scores for 337,208 unrelated British European individuals from the UK Biobank. We then tested for interactions between the PGS and 20+ environmental factors including sex, diet, smoking, sedentary behavior, and socioeconomic status, identifying 15 multiplicative and 73 additive (FDR < 0.05) interactions. For instance, we find that a PGS stratifies T2D prevalence by 67% more within smokers compared to non-smokers, as measured by the difference in T2D prevalence between the top and bottom deciles of the T2D PGS distribution. Finally, we extended our approach to identify 61 additive interactions between the PGS, sex, and an environmental factor, which we define as sex-dependent GxE interactions (SxGxE). For example, we find that a PGS stratifies T2D rates by 100% more within females who recorded high sedentary behavior levels (top 20%) compared to those who recorded low levels (bottom 20%), as opposed to 120% more in males, indicating that sex influences the degree to which sedentary behavior modulates genome-wide contribution to T2D.
Overall, we demonstrate that the aggregation of genetic effects through PGS is a promising and powerful approach for detecting polygenic GxE. Using our method, we found that two-way and three-way GxE may play an important role in the architecture and prediction of human complex traits. However, the power of our approach is limited by sample size and the prediction accuracy of PGS, further motivating the inclusion of more diverse cohorts in future genetic studies.

S50. The current environment for gene-environment interactions

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 321. Gene-environment interaction between *TRIP4 l*ocus and air pollution exposure influences risk of coronary artery disease

Authors:

J. Hartiala[1], J. R. Hilser[1,2], Y. Han[1,2], Z. Cai[1], S. Biswas[1,3], F. D. Gilliland[1], H. Hu[1], J. A. Araujo[4,5,6], L. Ma[7], C. Miller[8,9,10,11], W. Mack[12], C. Finch[13], J. Bjorkegren[7], H. Allayee[1,2]; [1]Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Dept. of Biochemistry & Molecular Med., Univ. of Southern California, Los Angeles, CA, [3]Dept. of Surgery, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [4]Div. of Cardiology, David Geffen Sch. of Med. at Univ. of California, Los Angeles, Los Angeles, CA, [5]Dept. of Environmental Hlth.Sci., Fielding Sch. of Publ. Hlth.at UCLA, Los Angeles, CA, [6]Molecular Toxicology InterDept.al Program, Univ. of California Los Angeles, Los Angeles, CA, [7]Icahn Sch. of Med. at Mount Sinai, New York, NY, [8]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [9]Dept. of BioMed. Engineering, Univ. of Virginia, Charlottesville, VA, [10]Dept. of Biochemistry & Molecular Genetics, Univ. of Virginia, Charlottesville, VA, [11]Dept. of Publ. Hlth.Sci., Univ. of Virginia, Charlottesville, VA, [12]Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [13]USC Davis Sch. of Gerontology, Univ. of Southern California, Los Angeles, CA

Abstract:

Objective: In the present study, we leveraged the extensive clinical, genetic, and traffic-related air pollution (TRAP) exposure data available in the UK Biobank to investigate gene-environment (GxE) interactions for coronary artery disease (CAD) with known susceptibility loci. **Gene-Air Pollution Interactions with Known CAD Loci:** Logistic regression was used to test the lead SNPs at 213 previously identified loci for CAD for GxE interactions with levels of fine particulate matter <2.5mM in diameter ($PM_{2.5}$). At the Bonferroni-corrected threshold for multiple comparisons ($p<0.05/213$), rs6494488 on chromosome 15 significantly modulated association between $PM_{2.5}$ and risk of CAD (p-interaction=$2.0x10^{-04}$). More specifically, a 2-SD increase in $PM_{2.5}$ levels increased risk of CAD by 28% among those carrying 2 copies of the A risk allele (OR=1.28, 95% CI 1.25-1.32, $p=3.7x10^{-58}$) and by 17% among individuals carrying only one copy (OR=1.17, 95% CI 1.11-1.23, $p=3.1x10^{-09}$). **Prioritization of *TRIP4* as a Positional Candidate Gene:** Data from eQTLGen, GTEx Project, and the STARNET cohort revealed expression quantitative trait loci (eQTLs) with rs6494488 in blood for several positional candidate genes at this locus. However, eQTLs were observed only for *TRIP4* in other CAD-relevant tissues, such as aorta, coronary artery, and visceral adipose. Notably, *TRIP4* mRNA levels were lower in carriers of the CAD risk allele (A) in all tissues for which an eQTL was identified, suggesting that *TRIP4* has a protective role in CAD. Consistent with this hypothesis, *TRIP4* expression was lower in atherosclerotic aortas of CAD patients compared to aortas from CAD-free subjects. Single-cell RNA-seq data further revealed *TRIP4* expression to be abundant in subsets of endothelial cells, smooth muscle cells, and monocytes/macrophages of atherosclerotic plaques. Lastly, mRNA levels of *TRIP4*, but no other positional candidate gene, were significantly downregulated in endothelial cells cultured with plasma obtained from subjects exposed to TRAP. **Conclusions:** We identified a novel GxE interaction with $PM_{2.5}$ levels and risk of CAD that could be mediated through decreased *TRIP4* expression in multiple vascular cells that play important roles in atherogenesis.

S50. The current environment for gene-environment interactions

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 322. Genome-wide gene-environment interaction scans reveal novel putative drug targets: an Alzheimer's disease study

Authors:

Y. Zhang[1], S. Li[1], G. Abdel-Azim[1], S. Guo[2], K. He[3], E. Khramtsova[1], G. Rajagopal[1], S. Lovestone[4], W. Gauderman[5], M. Black[6]; [1]Janssen, Spring House, PA, [2]Janssen R&D, Johnson & Johnson, North Wales, PA, [3]Janssen R&D, San Mateo, CA, [4]Janssen, Oxford, United Kingdom, [5]Univ. of Southern California, Los Angeles, CA, [6]Janssen Res. & Dev. (Johnson & Johnson), Lansdale, PA

Abstract:

Objective: Gene-environment (GxE) interactions can indicate novel targets and inform patient stratification for drug development. Detailed phenotypic and genomic sequencing data available in the UK Biobank (UKB) enable comprehensive exposure- and genome-wide GxE analyses. We developed an analytical workflow for biobank-scale data and performed genome-wide GxE scans to elucidate targets for Alzheimer's disease (AD).
Methods: AD cases defined by diagnosis codes were compared with age- and sex-matched controls without medical history of any mental, behavioral, or nervous system disorders. We assembled 59 exposures assessed at UKB enrollment from multiple domains, including physical activity, lifestyle, sleep, as well as environmental and sociodemographic factors ('E factors'). We assessed marginal gene, GxE, and joint gene-GxE effects on AD in unrelated UKB participants of European decent with imputed array and whole exome sequencing (WES) data. Gene-based GxE burden tests across 15 models collapsing putative loss-of-function (LOF) and/or deleterious variants were assessed using WES data.
Results: Genome-wide variant-level marginal models replicated 4 known (*BIN1*, *MS4A*, *APOE* region and *PILRA*) and 3 new AD risk loci. Joint gene-GxE models revealed 12 additional loci significantly associated with AD, none of which would have been discovered by standard GWAS alone. Significant interactions ($p<5e-8$) with sleep, physical activity, or smoking contributed to their discovery. Half of these loci showed significant interactions with sleep. For example, *ASIC2* rs317396 A-allele (intronic, MAF=0.254) was associated with decreased AD risk in those who sleep more hours per day. Compared with the G/G homozygotes, AD risk was decreased for A-allele carriers who sleep >9 hours/day (OR = 0.69 for heterozygotes, 0.48 for homozygotes) but increased in those who sleep <5 hours/day (OR = 1.54 for heterozygotes, 2.42 for homozygotes). Exome-wide gene-level marginal models identified one known (*ABCA7*) and 2 novel AD genes; joint gene-GxE models yielded 44 additional risk genes, such as *ABHD17A*, for which greater physical activity mitigated the deleterious effect of LOF on AD risk.
Conclusion: These results facilitate the study of disease pathogenesis and point to novel putative drug targets for AD that were not previously detected by standard GWAS. Significant GxE interactions can also indicate patients most likely to benefit from targeted therapy. Our analytical workflow may be applied systematically to a wide array of diseases captured in biobank-scale datasets to drive novel target discovery and inform precision drug development.

S50. The current environment for gene-environment interactions

Location: Conv Ctr/Petree C/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 323. Evidence of gene x environment interactions acting on body mass index using genotyping and whole exome sequencing data

Authors:

A. Dawes, G. Hawkes, R. Beaumont, M. Weedon, A. Wood, T. Frayling; Univ. of Exeter, Exeter, United Kingdom

Abstract:

Genetic studies have identified 1000s of genotype-phenotype associations, but very few robust examples of gene-environment interactions. Some studies have shown that genetic risk scores for some diseases do not interact with lifestyle factors. An exception is genetic risk scores for body mass index (BMI), which have stronger effects in people at higher environmental risk. However, few studies have tested interactions with individual variants or genes, including rare variants, and there is no robust method to assess potential confounding of the environmental factor. We selected 76 common genetic variants associated with main effects on BMI, and 22 potential interacting risk factors, including age, sex and accelerometer-based measures of sleep and activity in up to 388,000 UK Biobank individuals. Where we found evidence of interaction, we assessed the potential for confounding by performing negative control tests - by repeatedly testing for interaction but using randomly sampled individuals of similar BMIs to those in the risk factor categories. We also replicated results for sex interactions using the GIANT consortium data and tested individual genes within relevant loci for evidence of interaction using whole-exome sequencing data and gene burden tests. We found evidence of interaction between, 9, 4, 3, 4, 4, 3, 1, 3, 2, 1, and 2 individual variants with age, sex, smoking, socio-economic position (SEP), self-reported activity, sedentary time, percentage fat intake, vegetable intake, mental health and wellbeing, self-reported chronotype and accelerometery-based measures of sleep timing, respectively, at FDR <0.05. These interactions included variants where the BMI associated allele had a stronger effect in females than males: G:rs543874 in the *SEC16B* locus($P_{interaction}$=1.67E-06), and A:rs9925964 in the *KAT8* locus ($P_{interaction}$=1.99E-05); a stronger effect in current smokers than non-smokers: A:rs1558902 at the *FTO* locus (P=7.01E-05), and A:rs7138803 at the *BCDIN3D* locus (P=2.30E-05); stronger effects in people of lower SEP: C:rs6567160 at the MC4R locus (P=2.52E-06), G:rs543874 at the *SEC16B* locus (P=2.48E-04) and C:rs10132280 at the *STXBP6* locus (P=4.24E-04). Variant-sex interactions were replicated in the GIANT data. Based on our negative control tests, 11 of these (FDR<0.05) putative interactions appeared specific to the interacting risk factor. We did not identify any evidence of interaction for genes in relevant loci using whole-exome sequence data and gene burden tests. Our data provide evidence for gene-environment interaction. Specific BMI variants have larger or smaller effects in people of different sex, age and activity levels.

S51. Using EHRs to withdraw new insights from biobanks

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 326. Variability in lifetime risk of 20 complex diseases across European countries and polygenic score strata in over 1 million individuals

Authors:

B. Wolford[1], B. Jermy[2], K. Läll[3], T. Hartonen[4], Z. Yang[2], R. Monti[5], Y. Wang[6], K. Zguro[7], O. Youssef[2], N. Mars[4], A. Martin[6], F. FinnGen[4], S. Furini[7], K. Hveem[8], H. Heyne[9], R. Magi[10], A. Palotie[4], A. Ganna[4], S. Ripatti[4]; [1]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [2]Univ. of Helsinki, Helsinki, Finland, [3]Estonian Genome Ctr., Tartu, Estonia, [4]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, [5]Hasso Plattner Inst., Univ. of Potsdam, Potsdam, Germany, [6]Massachusetts Gen. Hosp., Boston, MA, [7]Univ. of Siena, Siena, Italy, [8]NTNU, Trondheim, Norway, [9]Hasso Plattner Inst., Digital Hlth.Ctr., Potsdam, Germany, [10]Univ. of Tartu, Tartu, Estonia

Abstract:

Background Polygenic scores (PGS) have been shown to stratify individuals based on their relative risks in many common diseases. However, clinical decision making is typically based on absolute risk estimates, which make it easier to determine the optimal age and threshold for intervention.
Methods We developed a novel approach to estimate country-specific absolute lifetime risk (cumulative incidence by age 80) across PGS strata using biobank-specific PGS associations and age-, sex-, and country-specific metrics from the Global Burden of Disease (GBD). Phenotypes were harmonized and PGS computed using MegaPRS in 4 biobanks participating in the **INTER**national consortium of integrati**VE** ge**N**omics pr**E**diction (INTERVENE)—UK Biobank, FinnGen, Estonian Biobank, and Trøndelag Health Study. This is a combined sample size of ~1.06 million with extensive follow-up (i.e. >50 years of electronic health records in FinnGen).
20 diseases with summary statistics from well-powered genome wide association studies and contributions to the GBD were selected. Hazard ratios (HR) were estimated between each disease and relevant PGS using Cox-proportional hazards regression.
Results In Finland, T2D showed the greatest relative differentiation in lifetime risk when comparing individuals in the top 1% of PGS (24.7% [95% CI=23.0-26.4%]) to average (40-60%) PGS (5.6% [4.6-6.9%]). T2D was also the disease with the greatest relative difference between top 1% and average PGS strata in Norway (22.3% [20.4-24.0%] vs 3.4% [1.3-8.8%]) and Estonia (13.2% [12.1-14.2%] vs 3.7% [2.5-5.3%]).
In Finland, knee osteoarthritis, asthma, and coronary heart disease had relative differences >10% between top 1% and average PGS strata. In Estonia and Norway, knee osteoarthritis and asthma were also in the top 4 diseases with large differences between PGS strata.
Average lifetime risk of T1D showed the greatest difference between the UK (5.8%) and Finland (10.7%). This was due to both a reduced baseline incidence of T1D in the UK and a reduced association with T1D PGS (HR per SD: UKB=2.06 [95% CI=1.93-2.21], FinnGen=2.39 [2.32-2.46]). This trend was considerably magnified when analyzing the top 1% of PGS (HR: UKB=6.47 [4.61-9.10], FinnGen=9.94 [8.61-11.48]). We are currently working to replicate these findings in 700K samples from additional INTERVENE biobanks.
Conclusions To implement PGS into health care settings, the between-country variability in both lifetime risk estimates and genetic effects must be accounted for. Here, we present a novel unified strategy to integrate PGS effects and lifetime risk estimates. We quantify variability using 20 diseases and multiple European countries.

S51. Using EHRs to withdraw new insights from biobanks

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 327. Neuroimaging-guided PheWAS using electronic health records from UK Biobank and eMERGE cohorts yield novel associations between brain imaging and disease outcomes

Authors:

Y. Veturi[1], L. Guare[1], I. Kullo[2], H. Hakonarson[3], P. Sleiman[4], D. Schaid[2], D. Velez Edwards[5], Q. FENG[5], W-Q. Wei[6], M. Ritchie[4]; [1]Univ. of Pennsylvania, State College, PA, [2]Mayo Clinic, Rochester, MN, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Vanderbilt Univ., Nashville, TN

Abstract:

Background: Several studies have identified genetic overlaps between neuroimaging phenotypes and complex human diseases (e.g., Alzheimer's disease, type II diabetes, stroke). What shared genetic mechanisms exist between neuroimaging phenotypes and the spectrum of diseases in electronic health records (EHR) by examining multi-omic data? **Objectives:** A comprehensive investigation of functional associations between neuroimaging phenotypes and diseases in EHR **across the phenome** can yield novel image-derived phenotype (IDP) biomarkers for complex diseases as well as provide pleiotropic gene targets for drug repurposing. **Methods:** We conducted a genome-wide association study on 2,124 IDPs (structural and diffusion MRI measuring cortical volume, cortical thickness, brain volumes, cortical grey-white contrast etc.) from UK Biobank (UKB) across 12.5M SNPs on 38K samples of European ancestry. Subsequently we conducted gene-based analyses using MAGMA as well as transcriptome-wide association study (TWAS) on 13 brain tissues from GTEx v8 followed by a gene-based colocalization protocol we developed. We also derived a polygenic priority score (POPS) for the detected genes by leveraging polygenic enrichments from gene expression, biological pathway and protein-protein interactions. We then mapped the significant genes to fine-mapped eQTL from PredictDB, obtained using multivariate adaptive shrinkage models. Finally, we conducted IDP-guided phenome-wide association study (PheWAS) on 12,494 fine-mapped eQTL across 664 ICD-10 codes (n=452,595 samples) in UKB and 424 ICD-9 codes (n=41,983 samples) in the electronic Medical Records and Genomics cohort (eMERGE). **Results:** We found 44 Bonferroni significant genes (15 chromosomes) from MAGMA ($p < 1.4E-9$) **and** TWAS ($p < 1.3E-10$) that passed coloc filters in brain tissues ($P[H3] < 0.5$ and $P[H4] > 0.9$) across 364 IDPs. 4 genes had a prediction score > 2, including known genes for white matter degeneration and brain atrophy (*PLEKHM1* and *EGFR*). IDP-guided PheWAS yielded 94 Bonferroni-significant eQTL in **both** cohorts ($p < 9E-09$) for 39 diseases in UKB and 20 diseases in eMERGE including neurological, immune-related, cardiovascular, endocrine and cancer. For example, of the 44 imaging-significant genes, eQTL mapping to *THBS3*, *BTN3A2*, *MSRA* were strongly associated with hyper/hypothyroidism, lupus erythematosus, hematuria and hyperplasia of prostate. **Conclusions:** We comprehensively examine shared functional associations between several neuroimaging measures and human diseases, with replication. Our analyses will help identify endophenotypes for future prevention and treatment strategies.

S51. Using EHRs to withdraw new insights from biobanks

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 328. Leveraging genomic diversity for discovery in an EHR-linked biobank: The UCLA ATLAS Community Health Initiative

Authors:

R. Johnson[1], Y. Ding[1], V. Venkateswaran[1], A. Bhattacharya[1], K. Boulier[1], A. Chiu[1], T. Schwarz[1], M. Freund[2], L. Zhan[1], K. Burch[1], C. Caggiano[1], B. Hill[1], N. Rakocz[1], B. Balliu[1], C. T. Denny[1], J. Sul[1], N. Zaitlen[1], V. Arboleda[1], E. Halperin[1], S. Sankararaman[1], M. Butte[1], UCLA Precision Health Data Discovery Repository Working Group, UCLA Precision Health ATLAS Working Group, C. Lajonchere[1], D. Geschwind[1], B. Pasaniuc[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]David Geffen Sch. of Med., UCLA, Los Angeles, CA

Abstract:

We explore the implications of genetic ancestry within the UCLA ATLAS Community Health Initiative - an ancestrally diverse biobank of genomic data linked with de-identified electronic health records (EHRs) of UCLA Health patients (N > 60K). We identified 5 continental-scale genetically inferred ancestry (GIA) clusters including European American (EA), African American (AA), Hispanic Latino American (HL), South Asian American (SAA), and East Asian American (EAA) individuals, as well as 7 subcontinental GIA clusters within the EAA GIA group including Chinese, Vietnamese, and Japanese American individuals. We observed marked differences between self-identified race/ethnicity (SIRE) and GIA where 13.33% of individuals within the AA cluster did not self-identify as Black/African American and 16.58% of the HL cluster did not identify as Hispanic/Latino. Analyzing >1,500 EHR-derived phenotypes (phecodes) within each GIA group, we found 732 phenotypes that showed cross-ancestry differences whose prevalence varied significantly by GIA. After accounting for SIRE, 259 significant associations remained, demonstrating that GIA provides information not captured by SIRE. Within the subcontinental EAA groups, we found meaningful disease associations not found when grouping individuals at the broader continental level. For example, no significant association for chronic kidney disease was found in the EAA group but was positively associated in the Filipino American group (OR: 1.83, p-value=2.87e-5) and negatively associated in the Chinese American group (OR: 0.54, p-value=2.90e-5). To demonstrate the utility of genetic data linked with EHR, we performed GWAS within the 4 largest GIA groups across 6 phenotypes spanning multiple disease categories and found 216 genome-wide significant SNPs. We observed ancestry-specific associations (e.g. skin cancer for EA), associations across multiple GIA groups (e.g. nonalcoholic liver disease for EA and HL), and 42 significant SNPs found only in meta-analyses computed across GIA groups but not within a specific GIA group. Lastly, we focused on liver disease, where GWAS identified significant associations in the 22q13.31 locus across the HL and EAA GIA groups. A subsequent PheWAS at the top SNP within both GIA groups revealed associations with multiple neurologic and neoplastic phenotypes that reached significance exclusively in the HL analysis, suggesting possible differential genetic architecture, as well as variation even at the phenotype level, reflecting possible genetic or environmental modifiers of important comorbidities.

S51. Using EHRs to withdraw new insights from biobanks

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Thursday, October 27, 2022, 1:45 pm - 2:45 pm

ProgNbr 329. Phenome-wide association study of polygenic risk for central adiposity

Authors:

N. Josyula[1], G. Chittoor[1], M. Graff[2], E. Wilson[2], K. North[2], W-Q. Wei[3], E. Marouli[4], G. Jarvik[5], K. Young[2], S. Berndt[6], C-T. Liu[7], R. Smit[8], I. Kullo[9], DiscovEHR collaboration, eMERGE Network, GIANT Consortium, K. Mohlke[10], A. Justice[1]; [1]Geisinger Hlth.System, Danville, PA, [2]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [3]Vanderbilt Univ., Nashville, TN, [4]Barts and The London Sch. of Med. and Dentistry Queen Mary Univ. of London, London, United Kingdom, [5]Univ. Washington Med. Ctr., Seattle, WA, [6]Natl. Cancer Inst., Rockville, MD, [7]Boston Univ. SPH, Boston, MA, [8]Icahn Sch. of Med. at Mount Sinai, New York, NY, [9]Mayo Clinic, Rochester, MN, [10]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Increased abdominal adiposity is a risk factor for several cardiometabolic and cardiovascular disorders independent of overall obesity. However, unlike body mass index (BMI), waist and hip circumference assessments are often not measured in regular preventive care. With the growing number of electronic health record (EHR)-linked Biobanks with available genotyping, but lacking central adiposity measures, we aim to assess the potential of using genetic risk for central adiposity using polygenic risk scores (PRS) for body fat distribution- measured as waist-to-hip ratio adjusted for BMI (WHRadjBMI) to identify elevated disease risk in the clinical setting. We hypothesized that a phenome-wide association (PheWAS) analysis of this PRS would validate expected central adiposity associations with cardiometabolic phenotypes and identify potentially novel associations with disease categories.

Using results from a trans-ancestry GWAS of adult WHRadjBMI from over 1,187,156 participants (80% European, 11% East Asian, 4% South Asian, 3% Hispanic/Latino, 2% African) in the GIANT consortium, we developed a PRS using PRS-CSx optimized in an independent study sample from the UK Biobank. The resulting SNPs and weights were used to evaluate the associations between the PRS and diseases/traits in an independent subset of 180,706 individuals from Geisinger's MyCode Study and the Electronic Medical Records and Genomics (eMERGE) Network. The analyses were adjusted for age, sex (for non-sex-specific codes), and principal components to control for ancestry. Study-specific PheWAS results were then meta-analyzed using inverse variance weighted fixed-effects meta-analysis.

Of the 1,820 PheCodes available in the meta-analysis, we identified 117 significant associations ($P<2.7e-05$), with top results validating expected central adiposity associations with cardiometabolic outcomes. In general, increasing PRS was positively associated with Type 2 diabetes (P=7e-187), hyperlipidemia (P=2e-72), coronary atherosclerosis (P=9e-53) and hypertension (P=9e-44). In addition to endocrine/metabolic and circulatory system phenotypes, we identified significant associations with behavioral disorders (i.e. tobacco use, P=2e-10), neurological (i.e. inflammatory and toxic neuropathy, P=6e-8), digestive (i.e., gastroesophageal reflux disease, P=6e-7) and respiratory (i.e. pulmonary collapse, P=2e-6) illnesses.

This analysis demonstrates the utility of applying central adiposity PRS for phenome-wide interrogation in identifying increased risk for important comorbidities in the EHR, even in the absence of clinically assessed body fat distribution.

S60. Computational approaches for understanding noncoding variants

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 378. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes improves the functional interpretation of non-coding regions

Authors:

S. Chen, L. Francioli, J. Goodrich, R. Collins, Q. Wang, J. Alfoldi, N. Watts, C. Vittal, L. Gauthier, T. Poterba, M. Wilson, Y. Tarasova, W. Phu, M. Yohannes, Z. Koenig, gnomAD Project Consortium, A. O'Donnell-Luria, M. Solomonson, C. Seed, A. Martin, M. Talkowski, H. Rehm, M. Daly, G. Tiao, B. Neale, D. MacArthur, K. Karczewski; Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract:

The depletion of disruptive variation caused by purifying natural selection (constraint) has been widely used to investigate protein-coding genes underlying human disorders, but attempts to assess constraint for non-protein-coding regions have proven more difficult. Several challenges arise when extending the gene constraint model to the non-coding space - in particular, the sample size of human whole-genome reference data has been relatively small compared to the exome; and the mutation rate in non-coding regions is highly heterogeneous, which can be affected not only by local sequence context as commonly modeled in gene constraint metrics but also by a variety of genomic features at larger scales.

Here we aggregate, process, and release a dataset of 76,156 human genomes from the Genome Aggregation Database (gnomAD), the largest public open-access human genome reference dataset, and use this dataset to build a mutational constraint map for the whole genome. We present an improved mutational model that incorporates local sequence context and regional genomic features to detect depletions of variation across the genome. We validated our metric using a series of external functional annotations, and by benchmarking our constraint metric against other genome-wide measures (e.g., Orion, CDTS, and gsRVIS), we show that our metric outperforms existing scores in identifying functional non-coding variants potentially associated with human diseases and traits.

Our analyses revealed significant correlation between constraint of the non-coding regulatory elements and the functional importance of their target genes. Interestingly, we also found a number of constrained regulatory elements being associated with "unconstrained" genes as classified by gene constraint metrics. We demonstrate that the lack of predicted gene constraint is attributed to the intrinsic design of gene scores measuring constraint against rare LoF variation, where small genes with few expected LoF variants are inevitably underpowered. This thus highlights the value of non-coding constraint - as a complementary metric to gene constraint - to improve gene function characterization. A practical implementation would be to incorporate the constraint of regulatory elements into the modeling of gene constraint, which essentially borrows power from extending the functional unit of a gene to encompass its regulatory components.

Together, we demonstrate that this genome-wide constraint map provides an effective approach for characterizing the non-coding genome and improving the identification and interpretation of functional human genetic variation.

S60. Computational approaches for understanding noncoding variants

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 379. Trans-PCO: a powerful approach to detecting *trans*-QTLs associated with regulatory networks

Authors:

L. Wang, X. Liu; Univ. of Chicago, Chicago, IL

Abstract:

Disruptions of regulatory networks is a key mechanism underlying human diseases. However, detecting *trans*-eQTLs of regulatory networks involving multiple genes remains challenging. While the first PC of multiple genes' expressions captures the largest expression variance, it can have little power in association tests (Liu 2019 JASA). To improve the power of detecting *trans*-eQTLs, we propose a PC-based multivariate association pipeline, trans-PCO, that combines multiple PCs to detect *trans*-eQTLs of regulatory networks. Our simulations demonstrated that trans-PCO substantially outperforms the existing primary PC approach and traditional univariate mapping: trans-PCO achieved 74% power at detecting *trans*-eQTLs, in comparison to 15% power of univariate associations and only 0.002% power of the primary PC method at the sample size of 800. We applied trans-PCO to two RNA-sequencing datasets in whole blood, DGN (N=913) and eQTLGen (N=31,684), to identify *trans*-eQTLs associated with gene co-expression networks. We carefully dealt with different sources of false positive *trans*-eQTLs. In total, we identified 3899 significant *trans*-eQTL SNP-module pairs in DGN, and 8199 significant *trans*-eQTL SNP-module pairs in eQTLGen at 10% FDR. We performed co-localization between *trans*-eQTLs and *cis*-eQTLs. We found that 45% *trans*-eQTL loci share causal variants with *cis*-eQTLs, suggesting a *cis*- mediated *trans*- regulatory mechanism. Yet, the remaining *trans*- loci are unexplained, which requires further study to understand their *trans*- mechanism. All DGN *trans*-eQTLs that overlap with eQTLGen SNPs are replicated in eQTLGen, indicating high replication rate across studies. To investigate the role of trans- regulation in complex trait genetics, we performed colocalization analyses of *trans*-eQTLs and GWAS loci of 29 blood traits and 10 autoimmune diseases. We observed high proportions (median=30.3%) of colocalization between *trans*-eQTL loci and blood traits. Interestingly, we found that *trans*-eQTLs of module 4 co-localized with 24 blood traits. Module 4 is highly enriched in a KEGG pathway: platelet activation (BH adjusted p-value=6.2×10^{-9}). We also estimated that module 4 is significantly enriched in the SNP heritability of mean platelet volume (Enrichment=6.68, P=1.22×10^{-5}) and platelet distribution width (Enrichment=6.49, P=7.03×10^{-5}). The trans-eQTLs of modules 4 reveal the *trans*-regulatory mechanisms of platelet related traits. In conclusion, trans-PCO is a very powerful and reliable tool that detects *trans*-eQTLs of cellular pathways and networks, which gives a more complete picture of the molecular basis of complex diseases.

S60. Computational approaches for understanding noncoding variants

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 380. Performing TWAS on proteomic data to understand *cis* and *trans* gene regulatory mechanisms underlying complex traits

Authors:

H. Wittich[1], S. Rich[2], J. Rotter[3], H. Im[4], A. Manichaikul[2], H. Wheeler[1], NHLBI TOPMed Consortium; [1]Loyola Univ. Chicago, Chicago, IL, [2]Univ. of Virginia, Charlottesville, VA, [3]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [4]Univ. of Chicago, Chicago, IL

Abstract:

Regulation of gene expression is an important mechanism through which genetic variation can affect complex traits. Much progress has been made in uncovering *cis*-acting expression quantitative trait loci (*cis*-eQTL), defined here as within 1 Mb of the transcription start site, but *trans*-eQTL have been more difficult to identify and replicate. Also, technological advances have made proteomics studies more feasible. Here we use our ability to predict the *cis* component of gene expression coupled with the gene mapping method PrediXcan to identify the protein targets of *cis*- and *trans*-acting genes.
We used 3,622 plasma proteins measured in 3,301 individuals from the European ancestry INTERVAL study for discovery and 1,305 plasma proteins measured in 971 individuals from the Trans-omics for Precision Medicine (TOPMed) Multi-Ethnic Study of Atherosclerosis (MESA) for replication. First, we predicted gene expression levels from genotype dosages using models trained in 49 tissues from the Genotype-Tissue Expression (GTEx) Project. As in PrediXcan, this step gives us an estimate of the genetic component of gene expression, GReX, for each gene in each tissue. In the second step, we tested each GReX estimate for association with the observed abundance of each protein measured via SOMAscan assay.
We observed 96,705 Bonferroni significant transcript-protein pairs across all 49 tissues, 37,259 of which were *cis*-acting, meaning the gene for the transcript and the gene for the protein were within 1Mb of each other, and 59,446 of which were *trans*-acting, meaning the gene for the transcript and protein were on different chromosomes or greater than 1 Mb away. 11,586 of these significant pairs replicated in TOPMed MESA. Correlation analysis of the effect sizes of significantly associated pairs showed that *cis*-acting relationships are more uniformly shared across the 49 GTEx tissues than *trans*-acting. Furthermore, when clustering around *cis*-acting effects, the brain tissues separate out into their own cluster, and whole blood and testis did not cluster well with any other tissues. As expected, gene set enrichment analysis found that the protein targets of significantly associated pairs were enriched for known plasma proteins. Furthermore, both *cis*- and *trans*-acting transcripts in significant pairs were enriched for plasma proteins and GWAS loci for autism spectrum disorder, ulcerative colitis, and inflammatory bowel disease. These results show that performing TWAS on proteome data reveals both *cis*- and *trans*-acting mechanisms underlying the genetics of complex traits.

S60. Computational approaches for understanding noncoding variants

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 381. Modeling tissue and gene co-regulation reveals causal tissues for disease

Authors:

T. Amariuta, K. M. Siewert-Rocks, A. L. Price; Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Abstract:

Identifying causal tissues for disease informs our understanding of disease etiology. While associated tissues have been identified for most diseases, co-regulation of gene expression across tissues and genes complicates the inference of causal tissues. Previous work emphasizes the potential of accounting for co-regulation across tissues (Ongen et al. 2017 Nat Genet), but does not formally model tissue-specific effects and does not consider co-regulation across genes. We developed Tissue co-regulation score regression (TCSC), a method that estimates tissue-specific contributions to disease. TCSC uses gene-disease associations across tissues from transcriptome-wide association studies (TWAS), which include direct effects of causal tissues/genes and tagging effects of co-regulated tissues/genes. TCSC regresses TWAS chi-square statistics on co-regulation scores quantifying correlations of *cis*-genetic components of gene expression across tissues and genes (due to shared causal variants or LD between causal variants). TCSC determines that a tissue is causal if genes with high co-regulation to that tissue have higher TWAS chi-square statistics than genes with low co-regulation to that tissue. TCSC can also identify tissue-specific contributions to the covariance between two diseases by analyzing products of TWAS z-scores. In simulations at realistic parameter settings, TCSC detects the causal tissue with 72% power, substantially higher than the Ongen et al. method, and properly controls type I error. We applied TCSC to gene expression data for 48 GTEx tissues and GWAS summary statistics for 99 diseases/traits (average N = 294K). TCSC identified 29 causal tissue-disease pairs at 10% FDR, including several biologically plausible novel findings: esophagus mucosa and eczema, left heart ventricle and platelet count, and esophagus muscularis and FEV1/FVC; for example, esophagus mucosa is a plausible causal tissue for eczema because acid reflux, a comorbidity of eczema, changes the mucosal immune profile, which may lead to atopic inflammation or eczema. Notably, TCSC implicated at most one causal tissue for most diseases/traits, whereas the Ongen et al. method often implicated several co-regulated tissues. We also applied TCSC to 626 pairs of diseases/traits with significant genetic correlation. TCSC identified 63 causal tissue-disease covariance pairs at 10% FDR. For the positive genetic covariance between eosinophil count and white blood cell count, whole blood contributed positive covariance while lymphocytes contributed negative covariance; this suggests that disease covariance may reflect distinct tissue-specific contributions.

S60. Computational approaches for understanding noncoding variants

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 382. Genetically predicted metabolite levels identify candidate pathways for common diseases

Authors:

F. Nyasimi[1], L. Yanyu[1], Y. Park[2], S. Lawal[3], A. Brown[4], A. Viñuela[3], E. Fauman[2], N. Palmer[5], H. Im[1]; [1]Univ. of Chicago, Chicago, IL, [2]Pfizer, Cambridge, MA, [3]Newcastle Univ., Newcastle, United Kingdom, [4]Univ. of Dundee, Dundee, United Kingdom, [5]Wake Forest Sch. of Med., Winston-Salem, NC

Abstract:

Metabolomic profiles capture the interaction of cellular processes and environmental exposures. They have the potential to improve our mechanistic understanding of disease and identify clinically relevant treatment targets. However, due to the high cost, metabolite profiling is limited to a relatively small set of individuals. Furthermore, reverse causality complicates interpretation because many observed metabolite differences are the result rather than the cause of the disease.
Here we propose a method (MetaboXcan) that leverages the power of large-scale genome-wide association studies (GWAS) and reference metabolite data to discover new mechanisms of disease etiology and validate existing ones. MetaboXcan tests the association between genetic predictors of metabolite levels and complex traits to nominate causal metabolic pathways. It can also be interpreted as a Mendelian Randomization approach where the genetically predicted metabolite is the instrument.
To train genetic predictors, we applied the PRS-CS method to the metabolite GWAS summary statistics from METSIM (n≈6000), the DIRECT consortium (n≈3000), UK Biobank biomarkers (n≈360K), and the Nightingale Study (n≈115K). We used individual level metabolite data from the Insulin Resistance Atherosclerosis Family Study (n≈192) as a validation set. The PRS-CS performs well with summary statistics and we observe an improvement of performance as sample size increases. As expected, predictive accuracy increases with heritability.
Association testing of predicted metabolite levels and complex disease traits produced a number of interesting results. For example, we found that the dopamine (p=1e-4) and glutamate (p=1e-3) pathways were among the top pathways associated with schizophrenia; these pathways are known targets of schizophrenia treatment. Adiponectin, currently investigated as a neuroprotective target in brain disorders, was associated with the glutamate pathway (p=1e-4). A number of positive controls lend more credibility to our results. As expected, we observed the creatinine sub-pathway was significantly associated with kidney disease (p=1e-11) and that fasting glucose was associated with the gluconeogenesis pathway. Our results highlight the potential of our approach to provide orthogonal lines of evidence to inform the biology of complex traits.

S60. Computational approaches for understanding noncoding variants

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 383. Multi-biofluid metabolome-wide association study illuminates the molecular basis in complex trait genetics

Authors:

K. Huang[1], J. Liu[2], G. Song[2], D. Panyard[3], Y. Wu[2], Y. Deming[4], J. Miao[2], Z. Zhao[2], C. Engelman[4], Q. Lu[2]; [1]Dept. of Statistics, Univ. of Wisconsin-Madison, Madison, WI, [2]Dept. of Biostatistics and Med. Informatics, Univ. of Wisconsin-Madison, Madison, WI, [3]Dept. of Genetics, Stanford Univ. Sch. of Med., Stanford, CA, [4]Dept. of Population Hlth.Sci., Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI

Abstract:

Genome-wide association studies (GWAS), in conjunction with functional genomic follow-up analysis and multi-omics data integration, have successfully linked non-coding variants to functional roles and provided insights into the genetic architecture of numerous human complex traits. In particular, pairing GWAS associations with expression quantitative trait loci (eQTL), researchers have pinpointed target genes and relevant biological tissues for many complex traits. However, a large proportion of trait heritability and GWAS hits remain unaccounted for, hinting the existence of alternative molecular mechanisms not tagged by transcriptomic information. Here, we present the largest multi-biofluid metabolome-wide association study (MWAS) to date. We benchmarked and optimized genetic prediction models for a total of 2,548 metabolites from cerebrospinal fluid, plasma, and urine, and performed MWAS for 530 complex traits in three biofluids using samples of European descent in UK Biobank (N=375,770). We found a total of 44,749 significant metabolite-trait associations under a false discovery rate (FDR) of 0.05. Among 530 complex traits, 233 traits exhibit significant associations with at least one metabolite. 127 out of 233 traits showed significant associations in multiple biofluids while 106 traits showed biofluid-specific associations. These associations provide critical new insights into the molecular basis of analyzed traits. For example, we found a strong and urine-specific association between androstenediol (3beta,17beta) disulfate (1) and cholelithiasis (P=5.7E-05), mostly driven by metabolite QTLs at the *SULT2A1* locus on chromosome 19 (lead SNP: rs62129966-A; P = 3.1E-19). *SULT2A1* encodes the enzyme dehydroepiandrosterone (DHEA) sulfotransferase, catalyzes the sulfation of DHEA, of which androstenediol is a direct metabolite in estrogen biosynthesis pathway, confirming the critical role of estrogen biosynthesis in cholelithiasis etiology. Taken together, joint analysis of metabolite QTL data and complex trait GWAS in multiple biofluids revealed a plethora of candidate metabolites that could mediate genetic effects on complex traits, and suggested new etiological mechanisms unexplained by commonly hypothesized transcriptomic regulation. This work improves our understanding of the molecular outcomes of common genetic variations and may have lasting impact on novel biomarker discovery, clinical diagnosis improvement, and therapeutics development.

S61. Extended applications of polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 386. The UK Biobank polygenic risk score release

Authors:

V. Plagnol[1], D. Thompson[1], D. Wells[1], S. Selzam[1], E-C. Yeh[2], F-J. Hsieh[2], C-H. Chen[2], R. Moore[1], I. Peneva[1], L. Alexander[1], K. Sharp[1], W. Tarran[1], B. Ed[1], F. Riveros-Mckay[1], R. Sivley[1], D. Palmer[1], P. Seth[1], J. Harrison[1], M. Futema[3], Genomics England Research Consortium, P-Y. Kwok[2], G. McVean[1], P. Donnelly[1], M. Weale[1]; [1]Genomics plc, Oxford, United Kingdom, [2]Inst. for BioMed. Sci., Academia Sinica, Taipei, Taiwan, [3]St George's, Univ. of London, London, United Kingdom

Abstract:

UK Biobank (UKB) provides a uniquely-powered cohort for enabling research into the uses and performance of polygenic risk scores (PRS) for a wide range of traits. However, the lack of a standardised set of PRS and benchmarks is a challenge for validating findings and comparing between different tools

To address this need, we have worked with UKB to provide to the research community individual level PRS data for 28 diseases and 25 quantitative traits across the entire cohort. When sufficient ancestry specific summary statistics were available, we used a cross-ancestry methodology to generate ancestry-specific PRS models. We further evaluated the PRS models in another UK-based cohort (the 100,000 Genomes - Genomics England - Project) and two East-Asian cohorts (Taiwan Biobank and the Taiwan Precision Medicine Initiative).

In individuals of European ancestry, performance (measured as odds ratio per sd of the PRS in a logistic regression adjusting for age and sex [OR/SD]) across disease traits ranged from 1.4 (ovarian cancer) to 3.9 (type 1 diabetes), with a median of 1.8. For coronary artery disease, and excluding statin users, the top 8% of the PRS was associated with a risk equivalent to those in familial hypercholesterolemia mutation carriers. Averaging across all diseases, the OR/SD of PRS was reduced by 9%, 14%, and 27% in individuals of South Asian, East Asian, and African ancestries respectively. For a given genetic ancestry, PRS results were consistent across biobanks, with similar performances between UKB and 100K Genomes Project (Pearson’s r of 0.95 for logOR/SD). The UKB PRS release outperformed a broad set of 76 published PRSs in multiple ancestries in UKB, with a significant improvement compared to the nearest competitor for 15 out of the 19 disease traits and 8 out of 9 quantitative traits. The European performance bias was partially mitigated by the cross ancestry methodology; for example, the type 2 diabetes PRS has equal predictive performance in UKB and in East-Asian cohorts.

To enable other researchers to reproduce reported metrics and perform evaluations of their own PRS against the UK Biobank PRS Release, we have made available an open source tool within the UK Biobank Research Access Platform, along with the associated phenotype definitions. To quantify the impact of additional training, an alternative PRS for each disease, referred to as “enhanced” and trained using the UK Biobank White British unrelated sample set, is also provided as part of the release.

By providing these data within the UKB research platform, we provide a reference performance point for research into PRS and a readily-available resource for applications that require optimised PRSs.

S61. Extended applications of polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 387. Can genetic associations for disease onset be used to predict disease prognosis? The glass is half empty

Authors:

Z. Yang[1], J. S. Wanner[2], **N. Mars**[1], Finngen, INTERVENE, S. Ripatti[1], H. Heyne[2], A. Ganna[1]; [1]Inst. for Molecular Med. Finland, Helsinki, Finland, [2]Hasso Plattner Inst., Digital Hlth.Ctr., Potsdam, Germany

Abstract:

With the emerging of large-scale biobanks, longitudinal data has become increasingly available, which results in a growing interest in uncovering the genetic basis of disease prognosis. A natural question is whether the genetic underpinnings of disease onset can also be used to predict its prognosis. To test this hypothesis, we generated polygenic risk scores (PRSs) from GWAS of disease onset for 10 diseases. We then evaluated the PRSs ability in predicting diseases incidence as well as their prognosis (disease-related mortality or progression to a disease-specific complication) using Cox proportional hazard in 392,649 FinnGen participants. Overall, contrary to their good performance on predicting the disease incidence, onset PRSs are worse predictors for disease prognostic outcomes. Rather, scores developed on a similar prognostic trait in the general population tend to work better. For example, a PRS for coronary heart disease (CHD) onset had lower association with subsequent CHD-related mortality than a PRS for longevity in the general population (HR=1.05, $p = 9.16 \times 10^{-3}$ vs HR=1.10, $p = 2.97 \times 10^{-8}$). We also found varying predictive value of incidence PRS for markers of disease prognosis across different clinical endpoints. For instance, we explored two prognostic outcomes for atrial fibrillation (AF): PRS generated from GWAS of AF weakly predicts stroke (HR = 1.07, $p = 1.07 \times 10^{-3}$), whereas it had a stronger association with receiving a cardiodefibrillator (HR = 1.13, $p = 5.34 \times 10^{-24}$). These observations will be further validated across a biobank network with 1.2 million samples. We investigated reasons behind such observations using both real and synthetic data. As expected, we found some commonality but also considerable degree of distinctions in SNP effect distributions comparing disease onset and prognosis GWAS. We further examined these changes in SNP effects through analytical modeling and simulating under various disease genetic architectures. *Our result showed that even for SNPs with shared effect on both disease onset and prognosis, an attenuation in effect sizes can be observed in the prognosis GWAS partially due to selection bias.* Our results provide insights into genetics of diseases prognosis and set the expectations for the upcoming wave of prognosis GWASs. Meanwhile, it serves as a starting point for subsequent development of clinically useful prognostic genetic scores.

S61. Extended applications of polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 388. An atlas of associations between plasma protein biomarkers and polygenic risk scores for complex human diseases

Authors:

D. Chasioti[1], J. W. Robinson[2], H. Runz[1], C-Y. Chen[1], P. C. Haycock[2], J. Yarmolinsky[2], T. R. Gaunt[2], B. B. Sun[1], C. D. Whelan[1]; [1]Biogen Inc., Cambridge, MA, [2]Univ. of Bristol, Bristol, United Kingdom

Abstract:

Progress towards the clinical application of polygenic risk scores (PRSs) as risk prediction tools has been slow. Broader use of PRS to interrogate protein networks driving causal mechanisms underlying complex illnesses is largely under-explored. Here, we systematically investigated associations between polygenic risk for 32 diseases and relative concentrations of 1,463 blood plasma proteins, measured across 35,571 white European participants in the UK Biobank (UKB). We calculated PRS using the Bayesian-based polygenic risk score continuous shrinkage (PRScs) method, sourcing summary statistics from GWAS excluding UKB participants. The 1000 Genomes European sample catalog was used as an external reference panel for LD structure. We performed linear regressions between protein concentrations and each of the 32 disease PRSs, adjusting for age and gender. For Alzheimer's disease (AD), the two *APOE* genotype-defining SNPs were excluded from PRScs and included as regression covariates. Genetic risk for cerebral artery disease was associated with 217 proteins, with the strongest association for *PCSK9* ($p=1.38x10^{-39}$); Type II diabetes PRS was associated with 250 proteins, showing the strongest with *GSTAT1* ($p=1.35x10^{-41}$); and chronic kidney disease was associated with 443 proteins, the strongest of which was with uromodulin ($p=4.1x10^{-63}$). PRS for breast, endometrial, and prostate cancers showed the strongest associations with butyrophilin proteins *BTN3A2* ($p<8.76x10^{-35}$) and several MHC-encoded proteins, such as *MICA* glycoprotein ($p<6.04x10^{-14}$). Breast and prostate cancer PRS also significantly associated with *BTN2A1* ($p<1.61x10^{-12}$). Notably, schizophrenia PRS showed similar associations with *BTN3A2* ($p=5.8x10^{-211}$), *BTN2A1* ($p=2.79x10^{-70}$) and *MICA* ($p=1.35x10^{-110}$), which remained significant after adjusting for smoking status ($p<1.35x10^{-71}$). Multiple sclerosis PRS associated with 70 proteins enriched for interactions between lymphoid and non-lymphoid cells and CNS immune surveillance, such as *LILRB4* ($p<1x10^{-314}$). PRS for neurodegenerative illnesses associated with smaller clusters of proteins; for example, AD PRS associated with the innate immunity proteins *sTREM2* ($p=2.43x10^{-8}$), *CD33* ($p=1.09x10^{-6}$), ribonuclease *T2* ($p=3.05x10^{-6}$) and complement *C1Q* ($p=3.63x10^{-7}$) whereas Parkinson's disease (PD) PRS associated with six metabolic proteins including *BST1* ($p=1.26x10^{-7}$). Our results provide insights into potential causal disease pathways and highlight candidate biomarkers for further exploration. Future analyses will integrate additional disease endophenotypes and explore the utility of protein PRS as stratification tools.

S61. Extended applications of polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 389. Selection, optimization, and validation of 10 polygenic risk scores for clinical implementation in diverse populations

Authors:

E. Kenny[1], N. Abul-Husn[1], O. Dikilitas[2], T. Edwards[3], T. Ge[4], H. Hakonarson[5], M. Irvin[6], C. Kachulis[7], E. Karlson[8], A. Khan[9], M. Lape[10], C. Liu[9], J. Linder[3], T. Manolio[11], A. Miller[2], E. Malolepsza[7], J. Mosley[3], B. Namjou[10], J. Peterson[3], E. Rosenthal[12], R. Rowley[11], P. Sleiman[5], M. Thomas[13], H. Tiwari[6], The eMERGE IV Network, N. Lennon[14]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Mayo Clinic, Rochester, MN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Univ. of Alabama at Birmingham Sch. of Publ. Hlth., Birmingham, AL, [7]Broad Inst. of MIT and Harvard, Boston, MA, [8]Brigham and Women's Hosp., Boston, MA, [9]Columbia Univ., New York, NY, [10]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [11]Natl. Human Genome Res. Inst., Bethesda, MD, [12]Univ. of Washington Sch. of Med., Seattle, WA, [13]Fred Hutchinson Cancer Ctr., Seattle, WA, [14]Broad Inst. Genomic Services, Cambridge, MA

Abstract:

Polygenic risk scores (PRS) for myriad common diseases have improved in their predictive performance supporting their potential use in clinical practice. However, reduced predictive performance of PRS in diverse populations may exacerbate existing health disparities in clinical care. The NHGRI-funded eMERGE IV Network will return a PRS-based genome-informed risk assessment to 25,000 diverse adults and children. To develop clinical PRS genetic tests for this program, we initially assessed PRS performance, medical actionability, and potential clinical utility for 22 conditions, and selected 12 to be brought forward for development and quantitative assessment. PRS for 5 of the selected conditions were developed or optimized using cross-ancestry approaches and best performing models were selected. All PRS were validated in independent multi-ancestry datasets. We determined that PRS passed validation if the genomic predictor was significantly discriminative in a minimum of two and up to four validation populations; African/African-American (AA), Asian (Asn), European Ancestry (EA); and Hispanic/Latino (HL). Seven other standardized metrics were also considered in the selection process, and additional consideration was given to the strength of evidence in AA and HL populations. Ten conditions were selected with a range of high risk thresholds (top 2%-10%); atrial fibrillation, breast cancer, chronic kidney disease, coronary heart disease, hypercholesterolemia, and prostate cancer (in adults only); asthma and type 1 diabetes (in children only); obesity and type 2 diabetes (in adults and children). The reported odds ratios in the high risk group varied; median 2.5 and range 1.5-20.5 (across 10 conditions) in AA, 3.8 and 2.2-5.7 (5) in Asn, 3.2 and 2.3-13 (10) in EA, and 2.3 and 1.9-6.9 (7) in HL. We developed an implementation pipeline for clinical PRS testing, including a method that uses genetic ancestry to calibrate PRS mean and variance, a framework for regulatory compliance, and the development of a PRS clinical report in both pdf and structured data formats. Our experiences in eMERGE inform our understanding of the infrastructure and resources needed to implement PRS-based genomic screening programs in distinct clinical settings and for diverse communities.

S61. Extended applications of polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 390. The contributions of rare inherited and polygenic risk to autism spectrum disorder in multiplex families

Authors:

M. Cirnigliaro[1], T. S. Chang[2], S. A. Arteaga[2], L. Pérez-Cano[3], E. K. Ruzzo[1,4], A. Gordon[2], L. Bicks[2], J-Y. Jung[5,6], J. K. Lowe[7,4], D. P. Wall[5,6], D. H. Geschwind[2,7,4,8]; [1]Dept. of Psychiatry and Biobehavioral Sci., Semel Inst., David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [2]Dept. of Neurology, David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [3]STALICLA DDS, World Trade Ctr., Barcelona, Spain, [4]Ctr. for Autism Res. and Treatment, Semel Inst., David Geffen Sch. of Med., UCLA, Los Angeles, CA, [5]Dept. of Pediatrics, Div. of Systems Med., Stanford Univ., Stanford, CA, [6]Dept. of BioMed. Data Sci., Stanford Univ., Stanford, CA, [7]Dept. of Psychiatry and Biobehavioral Sci., Semel Inst., David Geffen Sch. of Med., UCLA, Los Angeles, CA, [8]Dept. of Human Genetics, David Geffen Sch. of Med., UCLA, Los Angeles, CA

Abstract:

The identification of rare *de novo* variants in autism spectrum disorder (ASD) has relied primarily on families with only one affected proband, leading to the identification of hundreds of risk genes harboring *de novo* variants. However, few studies have focused on the contribution of common or rare inherited variation to ASD genetic risk in families with more than one affected (multiplex families), despite the substantial heritability of ASD. Even less is known about the interplay of common and rare *inherited* variation. Previously, we found that multiplex families showed a different genetic architecture than simplex families, identifying several new risk genes with strong contributions from rare inherited variants and less from *de novo* risk variants. Here, we performed whole genome sequencing on the largest cohort of ASD multiplex families to date from the AGRE cohort, encompassing 4,551 individuals from 1,004 families, twice our previous study. Combining variants in this cohort with those from other studies, we identified 74 risk genes, including 7 new genes (*PLEKHA8*, *PRR25*, *FBXL13*, *VPS54*, *SLFN5*, *SNCAIP*, *TGM1*), mainly supported by rare inherited variation. We used functional genomic data, including single cell and bulk transcriptomes, to characterize the biological impact of these variants. We observed several notable complex patterns of inheritance. First, we observed six ASD children from three families inheriting rare variants in two known autism risk genes, consistent with an oligogenic burden in these families. Second, we confirmed the over transmission of ASD common genetic risk from parents without ASD to their affected children. Additionally, we observed this signal in ASD children carrying rare inherited variants in risk genes, not those with *de novo* variation, consistent with a mixed burden of rare and common inherited variation in affected children. Stratifying children with ASD based on their language, we observed over transmission of ASD polygenic risk (PGR) only in those with delayed language development, indicating that language, rather than being considered a peripheral phenotype, should be considered a core biological feature of ASD. This work also reveals a prominent role for both rare and common inherited variation in ASD risk, suggesting an additive effect that can explain why parents transmitting rare variants are unaffected, as affected children inherit both rare variants *and* PGR from both parents.

S61. Extended applications of polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 391. Germline polygenic risk scores predict cancer patient response and treatment failure

Authors:

D. Kurant, S. Groha, K. Kehl, A. Gusev; Dana-Farber Cancer Inst., Boston, MA

Abstract:

Background: Predicting which cancer patients are unlikely to respond to treatments is a fundamental goal of precision oncology and may greatly improve patient outcomes. Germline Polygenic Risk Scores (PRS) are an estimate of an individual's genetic risk for a trait/disease that have gained traction in stratifying patients for cancer risk, but have largely not been investigated in the setting of treatment response. Here, we sought to evaluate whether PRS can predict clinical meaningful cancer outcomes including treatment failure.
Methods: We evaluated 163,567 radiology reports from 16,182 unique cancer patients who had both radiological imaging, targeted tumor sequencing, and germline imputation as part of routine clinical care at the Dana-Farber Cancer Institute. PRS were computed for each patient for 220 common traits from public GWAS data, including cancer risk, autoimmune diseases, cardiovascular, and metabolic phenotypes. Natural language processing (NLP) was applied to the radiology reports to identify indicators of (i) patient mortality and (ii) treatment failure. As each NLP outcome is a longitudinal quantitative measure, we investigated multiple summary statistics of these outcomes for each patient, including: mean, median, min, max, and standard deviation. The association between PRS and summary NLP outcome were assessed by linear regression with gender, age, panel version, and cancer type as covariates, followed by multiple test correction.
Results: 12 out of 220 PRS were found to be significantly associated with patient mortality, and 10 were found to be significantly associated with treatment failure ($p<0.05$ after Bonferroni correction). Notably, PRS for breast cancer risk and lung cancer risk were highly significant and remained significant after excluding breast and lung cancer patients from the analysis, suggesting that these risk scores may capture broader pan-cancer processes.
Conclusions: We demonstrate that germline PRS correlate with treatment outcomes in a large cohort of cancer patients treated at a single tertiary institution. We propose that PRS has the potential to anticipate treatment change and mortality in cancer patients and may aide in optimizing populations for clinical trials.

S62. Genetics and integrated -omics of diabetes and associated metabolic disorders in diverse populations

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 394. Sequencing of 448 Greenlandic Inuit reveals a unique genetic architecture exemplified by a common high-impact *HNF1A* variant affecting type 2 diabetes

Authors:

F. Stæger[1], A. C. B. Thuesen[2,3], A. Kaciv[4,5,6], M. H. Solheim[4,7], I. Aukrust[8,4], E. Jørsboe[2], C. G. Santander[1], M. K. Andersen[2], Z. Li[1], A. Gilly[9], S. E. Stinson[2], A. P. Gjesing[2], P. Bjerregaard[10], M. L. Pedersen[11,12], C. V. L. Larsen[10], N. Grarup[2], M. E. Jørgensen[3,10,11], E. Zeggini[9,13], L. Bjørkhaug[14], A. Albrechtsen[1], I. Moltke[1], T. Hansen[2]; [1]Section for Computational and RNA Biology, Dept. of Biology, Univ. of Copenhagen, Copenhagen, Denmark, [2]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, [3]Clinical Res., Copenhagen Univ. Hosp. – Steno Diabetes Ctr. Copenhagen, Herlev, Denmark, [4]Ctr. for Diabetes Res., Dept. of Clinical Sci., Univ. of Bergen, Bergen, Norway, [5]Dept. of Pediatrics and Adolescents, Haukeland Univ. Hosp., Bergen, Norway, [6]Ctr. for Lab. Med., Østfold Hosp. Trust, Kalnes, Norway, [7]Dept. of Neuronal Control of Metabolism, Max Planck Inst. for Metabolism Res., Cologne, Germany, [8]Dept. of Med. Genetics, Haukeland Univ. Hosp., Bergen, Norway, [9]Inst. of Translational Genomics, Helmholtz Zentrum München – German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, [10]Ctr. for Publ. Hlth.Res. in Greenland, Natl. Inst. of Publ. Hlth., Univ. of Southern Denmark, Copenhagen, Denmark, [11]Steno Diabetes Ctr. Greenland, Nuuk, Greenland, [12]Greenland Ctr. for Hlth.Res., Inst. of Nursing and Hlth.Sci., Univ. of Greenland, Nuuk, Greenland, [13]Technical Univ. of Munich (TUM) and Klinikum Rechts der Isar, TUM Sch. of Med., Munich, Germany, [14]Dept. of Safety, Chemistry and BioMed. laboratory Sci., Western Norway Univ. of Applied Sci., Bergen, Norway

Abstract:

From sparse genetic data we have previously identified several private but common disease variants in the historically small and isolated Greenlandic Inuit population. Based on this and the demographic history of the population, we hypothesize that the Greenlandic genetic architecture, affecting common diseases like diabetes, is markedly different from other populations. We here present unpublished whole genome data from 448 Greenlanders and subsequent genotyping in 4497 individuals.
We show that a large number of common variants are private to Inuit while only a low number of variants are rare. This is also true for predicted functionally important variants such as loss-of-function (LOF) with more than half of them being 'common' (MAF>5%). In fact, most genes are dominated by a single common LOF variant in stark contrast to other well-studied populations where the LOF burden of a gene consists of many ultra-rare variants.
To exemplify the difference, we performed a screen for missense and LOF variants in the 14 known maturity-onset diabetes of the young (MODY) genes. We found one rare known pathogenic variant and one novel common variant (c.1108G>T). Using a mini-gene assay we show that the novel variant affects the splicing of *HNF1A* and a local ancestry analysis reveals that the variant is Inuit-specific with an Inuit allele frequency of 1.9%. Using a mixed model approach that takes ancestry and relatedness into account, we find that the variant is strongly associated with type 2 diabetes (OR=4.35, p-value=7.24×10^{-6}) in 4497 Greenlandic individuals. The variant is also associated with lower 30-min serum insulin (β=-232 pmol/L, β_{SD}=-0.695, P=$4{\cdot}43\times10^{-4}$) and higher 30-min plasma glucose (β=1.20 mmol/L, β_{SD}=0.441, P=0.0271) during an oral glucose tolerance test.
We found more carriers of this single variant (n=119) among Greenlanders, than carriers of any *HNF1A* LOF variant (n=109) across 36 different variants in 386,145 individuals of the UK biobank. However, in contrast to UK biobank, we only find a small number of Greenlandic carriers of variants in any of the other MODY genes. Due to the combination of large effect size and high allele frequency, the *HNF1A* variant has a large population impact on diabetes explaining 2.5% of the liability-scale variance (LVE). This is more than twice the effect of the largest population impact variant found in Europe or Asia (*TCF7L2*: 1.1% LVE). Together with the recessive *TBC1D4* variant, we show that 18% of diabetics in Greenland have a suggested monogenic diabetes etiology compared to 1-3% in large populations.

S62. Genetics and integrated -omics of diabetes and associated metabolic disorders in diverse populations

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 395. Trans-ancestry GWAS meta-analysis of random glucose provides insights into diabetes pathophysiology, complications, and treatment stratification

Authors:

M. Kaakinen[1,2], V. Lagou[3], L. Jiang[4], A. Ulrich[1], L. Zudina[1], Z. Balkhiyarova[1], A. Faggian[1], J. Maina[5], A. Demirkan[1], C. Reynolds[6], B. Jones[2], I. Prokopenko[1,5], The MAGIC Investigators; [1]Univ. of Surrey, Guildford, United Kingdom, [2]Imperial Coll. London, London, United Kingdom, [3]Wellcome Sanger Inst., Cambridge, United Kingdom, [4]The Univ. of Queensland, Brisbane, Australia, [5]Univ. of Lille, Lille, France, [6]Coventry Univ., Coventry, United Kingdom

Abstract:

Conventional measurements of fasting or postprandial blood glucose levels investigated in genome-wide association studies (GWAS) cannot capture the effects of DNA variability on the “around the clock” glucoregulatory processes. We performed GWAS meta-analysis of glucose measurements under non-standardised conditions using random glucose (RG) in 493,036 individuals of diverse ancestries and without diabetes, enabling powerful locus discovery and innovative pathophysiological observations. We analysed associations between HRC-imputed SNPs and RG, adjusted for age, sex, population structure, time since last meal (where available) in 17 studies, including UK Biobank. We additionally investigated genetic and causal relationships with other phenotypes through genetic correlation analysis, PRSs, hierarchical clustering and Mendelian Randomization, and gene expression through metaXscan and DEPICT. We discovered 139 RG loci (184 distinct signals), including 69 novel signals for glycaemia, 14 with sex-dimorphic effects, 9 identified through trans-ethnic analysis and 25 rare/low-frequency signals. Regulatory, glycosylation, and metagenomic annotations highlighted ileum and colon tissues, indicating an underappreciated role of gastrointestinal tract in the control of blood glucose. Functional follow-up and molecular dynamics simulations of low-frequency coding variants in *GLP1R*, a type 2 diabetes (T2D) treatment target, revealed that optimal selection of GLP-1R agonist therapy in the clinic will benefit from a tailored genetic stratification. We provide novel compelling evidence from Mendelian Randomization that lung function is modulated by blood glucose levels ($\beta_{MR\text{-}RG}$=-0.61, P=3.5x10^{-4}; $\beta_{MR\text{-}T2D}$=-0.062, P=1.42x10^{-21}), and settle the longstanding controversy that pulmonary dysfunction is a diabetes complication. Our investigation yields wide-ranging insights into the biology of glucose regulation, diabetes complications and pathways for treatment stratification. Funding: H2020-SC1-HBC-28-2019-LONGITOOLS, WCRF-2017/1641, Diabetes UK (BDA number:20/0006307), PreciDIAB (ANR-18-IBHU-0001).

S62. Genetics and integrated -omics of diabetes and associated metabolic disorders in diverse populations

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 396. Integrating transcriptomics, metabolomics, and GWAS helps reveal molecular mechanisms for metabolite levels and disease risk

Authors:

X. Yin[1], D. Bose[1], A. Kwon[1], S. Hanks[1], A. Jackson[1], H. Stringham[1], R. Welch[1], A. Oravilahti[2], L. Fernandes Silva[2], FinnGen, A. Locke[3], C. Fuchsberger[4], S. Service[5], M. Erdos[6], L. Bonnycastle[6], J. Kuusisto[2], N. Stitziel[3], I. Hall[7], J. Morrison[1], S. Ripatti[8], A. Palotie[8], N. Freimer[5], F. Collins[6], K. Mohlke[9], L. Scott[1], E. Fauman[10], C. Burant[1], M. Boehnke[1], M. Laakso[2], X. Wen[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Eastern Finland, Kuopio, Finland, [3]Washington Univ. Sch. of Med., St. Louis, MO, [4]EURAC, Bolzano, Italy, [5]Univ California los Angeles, Los Angeles, CA, [6]NHGRI/NIH, Bethesda, MD, [7]Yale Univ., New Haven, CT, [8]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [9]Univ. of North Carolina, Chapel Hill, NC, [10]Pfizer, Cambridge, MA

Abstract:

Transcriptomics data have been widely integrated with genome-wide association studies (GWAS) to help understand disease/trait molecular mechanisms. Less is known about how metabolites, such as lipids or components of metabolic pathways, link gene expression to disease processes. Our goal is to integrate metabolomics with transcriptomics and disease GWAS and identify metabolic pathways that underlie disease through identification of genetically co-regulated gene expression-metabolite pairs or gene expression-metabolite-disease triads. We performed probabilistic transcriptome-wide association and pairwise locus-level colocalization analyses to integrate transcriptomics results for 49 tissues in 706 individuals from the GTEx project, metabolomics results for 1,391 plasma metabolites in 6,136 Finnish men from the METSIM study, and GWAS for 2,861 disease traits in 260,405 Finnish individuals from the FinnGen study. We found that genetic variants that regulate metabolite levels were more likely to influence gene expression and disease risk compared to the variants that do not. Integrating transcriptomics with metabolomics results identified 397 potentially causal genes for 521 metabolites, of which 33% shared the same causal variants with genetic associations of gene expression. Integrating transcriptomics and metabolomics individually with FinnGen GWAS results identified 1,597 putative causal genes for 790 disease traits and identified metabolic pathways that may underlie disease. For example, we identified causal effects of *UGT1A1/UGT1A4* on gallbladder disorders through regulating plasma (E,E)-bilirubin levels, of *SLC22A5* on nasal polyps and plasma carnitine levels through distinct pathways, and of *LIPC* on age-related macular degeneration through glycerophospholipid metabolic pathways. Our study highlights the power of integrating multiple sets of molecular traits and GWAS results to deepen understanding of metabolic pathways underlying disease pathophysiology.

S62. Genetics and integrated -omics of diabetes and associated metabolic disorders in diverse populations

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 397. Integrated ca/eQTL analyses uncover diabetes GWAS variants modulating human islet proinflammatory cytokine responses

Authors:

R. Bhuiyan[1,2], R. Kursawe[1], C. N. Spracklen[3], D. Ucar[1,2], M. L. Stitzel[1,2]; [1]The Jackson Lab., Farmington, CT, [2]Univ. of Connecticut Hlth.Ctr., Farmington, CT, [3]Univ. of Massachusetts, Amherst, MA

Abstract:

Type 2 diabetes (T2D) is a complex genetic disease with significant environmental contributors that results from dysfunction and death of pancreatic islet cells. Genome-wide association studies (GWAS) identified variants representing >750 signals in >600 loci associated with T2D risk or progression (T2D variants). T2D variants are enriched in human pancreatic islet *cis*-regulatory elements (CREs), but only ~20% of these variants have been linked to steady-state alterations in CRE use or target gene expression. Pro-inflammatory cytokines (e.g. IL-1β/IFN-γ) contribute to islet dysfunction in T2D pathophysiology, but how genetic variants, especially T2D variants, modulate islet inflammatory cytokine responses is not yet understood.
We hypothesize that a subset of T2D variants alter islet inflammatory-responsive CRE use or activity and target gene expression to contribute to islet dysfunction and T2D. To test this hypothesis, we mapped chromatin accessibility and expression quantitative trait loci (ca/eQTL) from ATAC-seq and RNA-seq profiles of islets obtained from 70 genotyped individuals and exposed to proinflammatory cytokines. We identified 3,651 caQTLs (FDR < 5%) in cytokine-exposed islets, 1,768 of which modulated accessibility of cytokine-responsive CREs. Twenty-three T2D variants colocalized with these cytokine-specific caQTLs, including variants at the *ANKH*, *FADS1*, and *SLC7A7* loci. We also identified 1,831 cytokine-specific eQTLs (FDR < 5%), 222 of which modulated expression of 131 cytokine-responsive genes (e.g. *IL32*). Four T2D variants colocalized with cytokine-specific eQTLs such as *PTGFRN*. Integration of ca/eQTL results revealed 63 variants that modulate both cytokine-specific CRE use and target gene expression. At an exemplary locus, the T2D 'T' risk allele of the *SLC12A8* intronic variant rs4679370 reduced cytokine-responsive CRE use and decreased expression of *SLC12A8,* which encodes a nicotinamide mononucleotide transporter that protects islets against inflammation-related islet cell dysfunction. Together, this study reveals the genetic programming of human islet inflammatory cytokine responses and suggests that 27 T2D variants contribute to islet dysfunction in T2D risk/progression by altering islet resilience.

S62. Genetics and integrated -omics of diabetes and associated metabolic disorders in diverse populations

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 398. New common genetic loci for gestational diabetes mellitus reveal a distinct genetic architecture from type 2 diabetes mellitus

Authors:

A. Elliott[1,2,3,4], M. Kurki[1,2], M. Pirinen[4,5,6], R. Walters[1,2,3], J. Goldstein[2], FinnGen Study, A. Palotie[1,2,3,4], M. Daly[1,2,3,4], E. Widen[4]; [1]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [2]Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA, [3]Harvard Med. Sch., Boston, MA, [4]Inst. for Molecular Med. Finland, Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland, [5]Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland, [6]Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland

Abstract:

Gestational diabetes mellitus (GDM) occurs in 7-9% of pregnancies and is related to an increased lifetime risk of Type 2 diabetes (T2D). Thus, it has been proposed that these diseases share a common etiology and a common genetic predisposition, with GDM being a window that reveals an underlying predisposition to later-onset T2D. Five GDM loci have been identified to date, of which 4 are shared with T2D. While this appears to support a hypothesis of shared etiology, no study has been sufficiently powered to directly assess the degree to which genetic risk is shared in GDM vs. T2D and whether any variants and biological pathways are specific to, or have outsized importance in, GDM.

We conducted the largest genome-wide association study of GDM to date in 12,332 cases and 131,109 parous female controls in the FinnGen Study. We identified 14 independent significant signals in 13 loci, including 9 novel loci. GDM and T2D were genetically correlated (rg=0.710, se= 0.056, p=6.8e-37) but the correlation was significantly less than 1 (p=1.2e-7). Six of these loci are not significantly associated with T2D in a previously published large meta-analysis nor in FinnGen, while the remainder are established T2D hits. We built a Bayesian algorithm to explore the relationship between GDM and T2D effects. The approach rejects the existence of a single, consistent relationship between GDM and T2D across loci, instead defining two distinct classes of significant variants in this scan. For one group of associations, the classifier indicates 8 of the 13 to be GDM-predominant SNP effects with an effect size roughly 3x greater in GDM than T2D, while the remainder show a pattern identical to T2D variants not significant in GDM, all of which demonstrate a consistent directional association with reduced effect size in GDM versus T2D.

The 8 GDM-predominant associations include a putatively causal missense mutation on the X-chromosome (*MAP3K15*) that protects from GDM but increases hypertension risk, a variant mapping to the 5'UTR of the estrogen receptor, *ESR1*, and variants at the *MTNR1B*, *PCSK1*, *GCKR*, *CMIP*, *SPC25*, and *NEDD1*. Loci with GDM-predominant effects mapped to genes related to islet cells, central glucose homeostasis, steroidogenesis, and placental expression. These results pave the way for an improved biological understanding of GDM pathophysiology and its role in the development and course of T2D. Our results suggest that the genetics of GDM risk falls into two categories - one part T2D polygenic risk and one part uniquely gestational contributors to disease.

S62. Genetics and integrated -omics of diabetes and associated metabolic disorders in diverse populations

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 399. Identification of paradoxical genetic metabolic effects in Europeans, South Asians and Africans using 450,000 exomes and 150,000 whole genomes

Authors:

G. Hawkes, K. Patel, A. R. Wood, S. Martin, R. N. Beaumont, J. Tyrrell, M. N. Weedon, T. M. Frayling; Coll. of Med. and Hlth., Exeter Univ., Exeter, United Kingdom

Abstract:

Aim
Some genetic variants are associated with apparently paradoxical effects on adiposity and metabolism: alleles associated with higher body fat can be associated with a favourable metabolic profile, and vice-versa. However, study of these variants has been limited to monogenic disease and common variants, and usually performed in people of European ancestry. We aimed to discover novel genes and variants that have a paradoxical effect on body fat and metabolic profiles in multiple ancestries, using whole genome and exome sequencing data.
Methods
We performed a whole-genome sequencing-based variant study, testing approximately 140,000,000 variants in 150,000 individuals, and analysed more than 18,000 annotated exome-sequenced genes in 450,000 individuals, both of either European, South Asian or African ancestries, on body-fat % and metabolic phenotypes (HDL-cholesterol and triglycerides). We performed four different analyses to discover novel genes and variants: (1) independent phenotype testing (2) multivariate phenotype testing (3) a novel inverted genomic model, where we performed association testing of probabilities derived from a logistic model trained on carriers of pathogenic PPARG variants, which is a known monogenic cause of lipodystrophy, and (4) discordancy between BMI and T2D associations. We then determined the overlap between the methods, and performed a co-localisation analysis across phenotypes.
Results
Using whole-genome sequencing-based single variant testing, we found 97 independent variants potentially associated with a paradoxical metabolic phenotype, 32 of which appeared in multiple analyses. For example, a common variant close (<300kb) to *VEGFA* decreased BF% (p = 7.00e-4), increased triglyceride levels (p = 1.49e-29), decreased HDL (p = 5.71e-27), and was strongly associated with both the multivariate phenotype (p = 8.35e-54) and *PPARG*-lipodystrophy phenotype (p = 4.38e-47). We found little evidence that whole-genome sequencing data improved co-localisation compared to GWAS array data. Using whole-exome gene-burden testing, we found 30 genes associated with a paradoxical metabolic phenotype, including *PPARG*. Eleven loci were identified using both whole exome and whole genome sequencing testing.
Conclusion
In conclusion, we have performed the largest multi-ancestry analysis using whole genome and whole exome sequencing on a complex metabolic phenotype. We have discovered putative novel genes and variants with paradoxical metabolic effects and improved understanding of the biological mechanisms of these effects, using novel methods which combine information from multiple phenotypes.

S63. Hearing and seeing the advancements in auditory and vision genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 402. *Chd7* and *Sox2* cooperate to regulate development of the inner ear

Authors:

D. M. Martin[1], J. Gao[2], J. M. Skidmore[2], J. Cimerman[2], K. E. Ritter[2], Y. Raphael[2]; [1]Michigan Med., Ann Arbor, MI, [2]Univ. of Michigan, Ann Arbor, MI

Abstract:

Hereditary hearing loss and balance disorders remain important clinical problems that have become easier to diagnose but have unfortunately not greatly benefitted from advances in molecular therapies. An important step in designing molecular therapies for hereditary hearing loss and balance disorders is understanding their underlying mechanisms. In this study, we explored genetic interactions between *Chd7* and *Sox2*, two genes that are both highly expressed in and critical for early inner ear development. *Chd7* encodes the ATP-dependent chromatin remodeler CHD7, affected in CHARGE syndrome, and *Sox2* encodes Sex determining Region Y-box 2 family transcription factor SOX2. In humans, single copy loss of either *CHD7* or *SOX2* disrupts development of multiple organs and tissues, including the brain, eye, ear, and heart. SOX2 and CHD7 have been shown to physically interact in neural stem cells, but their potential interaction in the developing ear has not been explored. *Sox2* and *Chd7* are highly expressed in neurosensory progenitors within the otic epithelium and regulate both sensory and non-sensory development of the inner ear. We found that *Chd7* and *Sox2* exhibit dynamic co-expression in the developing mouse ear and that CHD7 acts genetically upstream of *Sox2*. We also found that double heterozygous ($Chd7^{Gt/+};Sox2^{CreER/+}$) mice exhibit early postnatal death and severely malformed inner ear vestibular and auditory structures, including abnormalities of the semicircular canals and shortened cochlea that may be explained by increased cell death. Through inducible deletions, we observed a critical window of *Chd7;Sox2* genetic interactions at E9.5. Differential gene expression analysis revealed 75 down-regulated genes and 56 up-regulated genes in $Chd7^{Gt/+};Sox2^{CreER/+}$ developing otocysts. The most highly upregulated gene was the LIM homeodomain transcription factor gene *Lhx1*, confirmed by immunostaining to be present in the developing otocyst. Interestingly, mice with single copy loss of both *Chd7* and *Lhx1* exhibited worsened cochlear hypoplasia than $Chd7^{Gt/+}$ mice, suggesting complex genetic interactions in the developing ear. These results demonstrate that expression of *Chd7* and *Sox2* are tightly regulated during inner ear development, and that interference with this expression leads to complex outcomes in downstream target gene expression that will inform novel targets for molecular therapy approaches to treat hearing and balance disorders.

S63. Hearing and seeing the advancements in auditory and vision genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 403. Syntaxin 4 is essential for hearing in human and zebrafish

Authors:

T. Bharadwaj Mathilakath, I. Schrauwen[1], A. Ghaffar[2], K. Shah[3], S. Rehman[2], A. Acharya[1], K. Liaqat[1], N. S. Lin[1], J. L. Everard[1], A. Khan[4], Z. M. Ahmed[2], W. Ahmad[5], S. Riazuddin[2], S. M. Leal[1,6]; [1]Columbia Univ., New York, NY, [2]Univ. of Maryland, Baltimore, MD, [3]COMSATS Univ. Islamabad, Abbottabad Campus, Abbottabad, Pakistan, [4]Hazara Univ. Mansehra, Khyber Pakhtunkhwa, Pakistan, [5]Quaid-i-Azam Univ. Islamabad, Pakistan, Islamabad, Pakistan, [6]Taub Inst. for Alzheimer's Disease and the Aging Brain, and the Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY

Abstract:

Congenital hearing impairment is a genetically highly heterogenous disorder in which prompt recognition and intervention are crucial to optimize outcomes. In this study, we investigated a large consanguineous Pakistani family with eight affected individuals showing bilateral severe-to-profound hearing impairment. Exome sequencing, homozygosity mapping and linkage analysis were used to determine the causative variant. We identified a homozygous splice region variant in STX4 (c.232+6T>C), predicted to impact splicing by computational tools, that segregated with hearing impairment (two-point LOD score = 5.9). STX4, a member of the syntaxin family, is a component of the SNARE machinery involved in several vesicle transport and recycling pathways. In-silico analysis of the expression of mouse orthologue Stx4a during mouse development were performed using various publicly available datasets with RNA-seq and/or microarray data of inner ear and craniofacial tissues. The analysis showed that murine orthologue Stx4a is highly and widespread expressed in the developing and adult inner ear, including the sensory epithelium, hair cells, spiral and vestibular ganglion cells, and is upregulated during development. Immunofluorescent imaging of mouse inner ear tissues revealed localization of STX4A in the cell body, cell membrane and stereocilia of inner and outer hair cells. To confirm its function in hearing and in the sensory epithelium, we next performed a knockdown of stx4 in zebrafish via a morpholino targeting the splice site donor corresponding with the human c.232+6T>C variant at the exon 3 - intron 3 boundary. This led to the in-frame retention of intron 3 and creation of premature stop codon via the novel inserted sequence in the mRNA. Startle response activity measured at 5 dpf showed a significant reduction in moderate activity following 1200Hz frequency sound between stx4-injected and un-injected larvae, suggesting impaired hearing. In addition, stx4-injected larvae also showed severe morphological and developmental defects with edema. Last, a disrupted mechanotransduction function in neuromast hair cells was seen via measuring FM1-43 uptake. In conclusion, our findings indicate that STX4 dysfunction leads to hearing impairment in humans and zebrafish and supports the evolutionary conserved role of STX4 in inner ear development and hair cell functioning.

S63. Hearing and seeing the advancements in auditory and vision genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 404. Whole genome sequencing for rare variants and imaging analyses in model organisms identify *SLC16A8* as a significant contributor for age-related macular degeneration risk

Authors:

N. Nouri[1], A. Stockwell[2], B. Hannon[1], T. Truong[1], Y. Malato[1], K. Browder[1], A. Sene[1], M. McCarthy[1], E. Chew[3], S. Chaney[1], H. Jasper[1], B. Yaspan[1]; [1]Genentech Inc., South San Francisco, CA, [2]Genentech Inc., Austin, TX, [3]NIH, Bethesda, MD

Abstract:

Polygenic predisposition to advanced age-related macular degeneration (AMD) enables strategies to define the complex pathogenesis of this disease. We combined human genetics and in vivo functional studies in *Drosophila melanogaster* to identify novel pathways implicated in AMD pathogenesis. Whole genome sequencing (WGS) and subsequent rare variant (RV) burden test analyses were performed for 6055 advanced AMD cases and 8294 non-AMD diseased controls. We extracted rare exonic SNPs (minor allele frequency <1%) predicted to alter amino acids or affect functionality of the protein. Since comparison cohorts were comprised of non-AMD clinical trial patients, a reverse regression with a Bayesian spike and slab prior was used to remove non-AMD specific associations. For functional characterization of candidate genes at association signals, we performed RNAi screens using *Drosophila* models. The impact of whole eye knockdowns of orthologous genes of interest on photoreceptor survival was assessed using deep pseudopupil (DPP) and optic neutralization of cornea imaging approaches that follow ommatidial morphology. WGS for RV burden tests pinpointed multiple genes, including the top hits *CFI* (*P*=3.13E-14; OR=4.57), *CFH* (*P*=2.14E-06; OR=2.05), and *SLC16A8* (*P*=1.50E-05; OR=1.78), all from loci previously associated with AMD (Table 1). We focused functional analysis on the proton-coupled monocarboxylate transporter *SLC16A8* (solute carrier family 16 member 8) which is expressed only in the retinal pigment epithelia of humans and also rodents. Knockdown of the *Drosophila* ortholog *sln* (silnoon) led to progressive loss of retinal homeostasis in the adult fly, as visualized by changes in photoreceptor EGFP expression patterns at day 14 (95% mean loss of DPP), but not by day 1, 7, or in comparison to negative controls (n=10 for each line and timepoint; reproduced in two separate assays). An unbiased, genome-wide analysis of rare coding variants provided additional support of a causal role for *SLC16A8* in protection against AMD when functional. We demonstrate that deletion of the *Drosophila* ortholog *sln* disrupts retinal homeostasis, and highlight a potential therapeutic opportunity for the treatment of AMD.

S63. Hearing and seeing the advancements in auditory and vision genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 405. Implication of a non-coding variation of *FOXE3* in an individual displaying a complex microphthalmia

Authors:

L. Fares Taie[1], C. Angée[1], J. Plaisancié[2], J-y. Douet[3], I. Raymond Letron[4], M. Goetz[5], C. Landrevie[5], M. Baltissen[6], C. Vincent-Delorme[7], M. Vermeulen[6], I. Karemaker[6], P. Calvas[8], N. Chassaing[9], J-M. Rozet[1]; [1]Imagine - Inst. of genetic diseases, Paris, France, [2]Hôpital Purpan, Toulouse, ToNIC (Toulouse Neuro Imaging Ctr.), Toulouse, France, [3]IHAP, Université de Toulouse, INRAE, ENVT, Toulouse, France, [4]Ecole Natl.e Vétérinaire de Toulouse, RESTORE Res. Ctr., Toulouse, France, [5]Université de Toulouse, ENVT, Toulouse, France, [6]Radboud Inst. for Molecular Life Sci., Oncode Inst., Radboud Univ. Nijmegen, Nijmegen, Netherlands, [7]Service de Génétique Clinique Guy Fontaine, Lille, France, [8]Hôpital Purpan, Ctr. de Référence des Anomalies Rares en Génétique Ophtalmologique (CARGO), Toulouse, France, [9]Hôpital Purpan, Toulouse, Service de Génétique Médicale, Toulouse, France, Toulouse, France

Abstract:

FOXE3 encodes a transcription factor (TF) which expression is extremely conserved throughout species that use vision. It plays a role mainly in the development of the lens, the integrity of which is required to allow proper development of the eye. Accordingly, biallelic truncating *FOXE3* mutations cause complex microphthalmia with lens, anterior segment and retinal anomalies. Studying an individual affected with a complex microphthalmia, we identified a monoallelic nonsense mutation in *FOXE3.* A search for a second hit in *trans* was undertaken leading to the identification of a previously unreported single nucleotide variant in an upstream conserved region 3kb distant from the gene. Here, we report the results of *in-vitro* and *in-vivo* genetically engineered mice studies designed to demonstrate the deleterious effect of the non-coding variant on FOXE3 function. Luciferase transactivation assay was performed which revealed a statistically significant impact of the non-coding variant on gene reporter expression as compared to the wildtype counterpart ($p<0.01$). This result led us to generate mouse lines carrying the non-coding variant or a nonsense mutation in homozygosity (KI/KI and KO/KO, respectively). The two lines were crossed to obtain compound heterozygous KO/KI animals as our patient. Phenotype analysis (Slit lamp, OCT, ocular size measurements) showed a spectrum of eye growth, cornea, anterior segment and lens abnormalities of increasing severity in the KI/KI, KI/KO and KO/KO lines, respectively. These anomalies are reminiscent of *FOXE3*-associated phenotypes in human. They were absent in wildtype and single heterozygous counterparts. To dissect the mechanisms by which the non-coding variant contribute to an ocular developmental defect, we performed DNA pull-down and mass spectrometry experiments using mouse ESC and immortalized human epithelial lens cells, in search for differentially bound TF. In total, we identified 35 candidate TF, three of which were common between the two cells lines. Interestingly, the knock-out of one of them has been reported to cause microphthalmia in the mouse. In conclusion, this work expands the mutational landscape of *FOXE3*-associated ocular developmental defects. Furthermore, it shows the contribution of unsolved cases to dissect the regulatory landscapes of genes and identify novel candidate genes.

S63. Hearing and seeing the advancements in auditory and vision genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 406. Beyond Mendelian Inheritance: An approach to digenic inheritance in cases of retinitis pigmentosa, deafness and Alport syndrome in a cohort of Colombian patients

Authors:

M. Galvez, J. Rincón Redondo, L. Rodríguez Vallejo, M. Latorre Quintana, S. Bello Uyaban; Gencell - Genuino Group, Bogotá, D.C, Colombia

Abstract:

Introduction: Digenic inheritance (DI) is the simplest form of oligogenic inheritance for complex diseases. Phenotypes where diagnostic suspicion includes a genetic component or are associated with incomplete penetrance, could be explained by a digenic model, as it was previously postulated. New genes and diseases with this inheritance pattern have been discovered and retinitis pigmentosa, deafness and Alport syndrome have been postulated as key models for digenic inheritance. **Objective:** To describe the possible combinations of genes associated with DI in a sample of Colombian patients who underwent whole-exome sequencing (WES) for retinitis pigmentosa, deafness, or Alport syndrome in whom a monogenic cause was not identified. **Methodology:** Observational, descriptive, retrospective cross-sectional study in a sample of patients with a suspected diagnosis of retinitis pigmentosa, deafness, or Alport syndrome, who underwent WES from 2019 to May 2022 and a monogenic cause was not identified. Variants reported as pathogenic, probably pathogenic, and of uncertain significance were considered in the analysis. Genes were searched in the Oligogenic diseases database (OLIDA) and the DigePred predictor validated by Vanderbilt University. The probable digenic combinations selected had a predictive value of $F_{0.5}$ >0,496, according to previously described by Mukherjee et al. **Results:** A total of 16 subjects were included (10 males), with an average age of 21,25 years. Analysis of possible digenic causal combinations included cases of deafness (68,75%), retinitis pigmentosa (18,75%), and Alport syndrome (12,50%). In 23,5% of the cases, one or several gene combinations with probable DI could be postulated. Additionally, seven new gene combinations and variants were identified, which included the genes *ABCA4, HGSNAT, USH2A, TUB, PRPH2,* and *ROM1* for retinitis pigmentosa, *ADGRV1* and *MYO7A* for deafness and *CEP290* and *TMEM67* for Alport syndrome. Of the combinations established by DigePred, 6 exceeded the established threshold, of which one (*PRPH2-ROM1*) had already been described in the literature. The other combinations are new findings. **Conclusions:** Looking for DI possibly increases the diagnostic rate of retinitis pigmentosa, deafness and Alport syndrome. New gene combinations of DI that support the diagnosis of these diseases have yet to be established, and further analysis are required. It is necessary to deepen the understanding of this type of inheritance.

S63. Hearing and seeing the advancements in auditory and vision genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 407. *FOXP4*: A novel candidate gene for angle closure glaucoma

Authors:

W. Presley[1], S. Q. Wang[1], E. Ward[1], K. Johnson[1], M. Shah[1], B. Guan[2], R. B. Hufnagel[3], J. E. Richards[1], L. Prasov[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]NIH, Bethesda, MD, [3]Natl. Eye Inst., NIH, Bethesda, MD

Abstract:

Primary angle closure glaucoma (PACG) describes a condition wherein anatomic blockage of the drainage angle of the eye leads to elevated eye pressure and optic neuropathy. PACG is present in 0.5% of the worldwide population, but the molecular pathogenesis of this condition is poorly understood, with few definitive PACG disease genes discovered to date. Plateau iris, a condition with anterior rotation of the peripheral iris, and high hyperopia can be associated with PACG. We identified a multigenerational pedigree with a triad of plateau iris, PACG and high hyperopia transmitted in an autosomal dominant fashion. Pooled whole exome sequencing revealed a missense substitution in a transcriptional repressor encoding gene, *FOXP4* c.1433 A>G:p.(Q478R), but no other compelling candidate variants. This variant has a gnomAD population allele frequency of only $6.8x10^{-5}$ and occurs at a highly conserved residue. *In silico* pathogenicity predictors (REVEL/MutScore/CADD) predicted it to be damaging, and homology modeling using SWISS-MODEL suggested that it is likely to interfere with protein dimerization. Spatiotemporal expression pattern of *FOXP4* was evaluated with immunostaining and RNAscope analysis in mouse embryonic eyes, and we found that it was highly expressed in multiple eye structures, including those relevant to drainage angle development, such as the periocular mesenchyme and cornea. Luciferase assays of the YFP-tagged FOXP4 p.Q478R variant in HEK-293T cells with the SPRX-Luc reporter showed that the variant protein retained its ability to repress transcription. Cellular localization analysis in HEK-293T cells showed that FOXP4 p.Q478R formed cytosolic aggregates not seen in the nuclear-localized wildtype protein, suggesting the variant allele may contribute to aggregation of the protein. Screening of 40 additional probands with PACG and/or high hyperopia did not identify any additional *FOXP4* variants overrepresented in cases over controls. Together, these continued analyses deepen our understanding of the role for *FOXP4* in both eye development and potential role in PACG.

S64. New molecular and analytical tools to study Mendelian disorders

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 410. PhenoScore: AI-based phenomics to classify genetic variants

Authors:

A. Dingemans[1,2], M. Hinne[2], K. M. G. Truijen[1], L. Goltstein[1], J. van Reeuwijk[1], N. de Leeuw[1], J. Schuurs-Hoeijmakers[1], R. Pfundt[1], I. Diets[1], J. den Hoed[3], E. De Boer[1], J. van der Spek[1], S. Jansen[4], B. van Bon[1], N. Jonis[1], C. Ockeloen[1], A. T. Vulto-van Silfhout[1], T. Kleefstra[1], D. A. Koolen[1], H. Van Esch[5], G. LYON[6], F. Alkuraya[7], A. Rauch[8], R. Marom[9], D. Baralle[10], P. J. van der Sluijs[11], G. Santen[11], F. Kooy[12], M. A. J. van Gerven[2], L. E. L. Vissers[1], B. B. A. de Vries[1]; [1]Radboud Univ. Med. center, Nijmegen, Netherlands, [2]Donders Inst. for Brain, Cognition and Behaviour, Nijmegen, Netherlands, [3]Max Planck Inst. for Psycholinguistics, Nijmegen, Netherlands, [4]Amsterdam UMC, Amsterdam, Netherlands, [5]Univ. Hosp. Leuven, Leuven, Netherlands, [6]Inst. for Basic Res. in Dev.al Disorders, Salt lake City, UT, [7]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [8]Univ. of Zurich, Schlieren-Zurich, Switzerland, [9]Baylor Coll. Med., Houston, TX, [10]Univ. of Southampton, Southampton, United Kingdom, [11]Leiden Univ. Med. center, Leiden, Netherlands, [12]Universtity of Antwerp, Edegem, Belgium

Abstract:

Algorithms using molecular information to predict the pathogenicity of genetic variants, such as the phyloP-, CADD- and SIFT score, are well established to classify genomic variants. An equivalent score to capture the phenotype is, however, not available, despite clinical features being one of the most significant predictors of pathogenicity of variants. We developed PhenoScore: an open source, artificial intelligence-based phenotypic pathogenicity score, combining facial recognition technology with phenotypic similarity in Human Phenotype Ontology (HPO). PhenoScore was estimated using facial images and HPO data after deep quantitative phenotyping of 501 individuals with 26 different rare neurodevelopmental disorders. We demonstrate the synergistic effect of this phenomics integration by featuring recognizable phenotypes for 25 out of 26 investigated genetic syndromes with a median area under the curve (AUC) of 0.87. Moreover, PhenoScore even works on small datasets, with data of only three individuals leading to a median AUC classification performance of 0.85. As expected, this improves when increasing the training dataset in size: to an AUC of 0.9 with seven individuals and 0.95 with 17 individuals in the training set. Moreover, this AI-framework PhenoScore was able to provide clinical evidence for two distinct *ADNP*-related phenotypes (p=0.01), that had already functionally been established but for which phenotype-driven analysis had so far failed to identify the respective clinical entities. Hence, PhenoScore will not only be of use to objectively quantify phenotypes to assist genomic variant interpretation, such as for reclassifying variants of unknown significance, but also allows for detailed genotype-phenotype studies to understand the molecular and phenotypic complexity of rare genetic syndromes. Furthermore, by investigating which features are important according to PhenoScore when making predictions, heatmaps for facial dysmorphisms and important clinical features can be visualized - enabling us to learn more about the phenotype of genetic syndromes as well. PhenoScore is open source and freely available to use and therefore easily applied to all other syndromes

S64. New molecular and analytical tools to study Mendelian disorders

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 411. A new locus for craniosynostosis caused by copy number variants flanking *FOXD3*

Authors:

A. Wilkie[1], E. Calpena[1], H. Mlcochova[1], A. Hashimoto[1], N. Koelling[2], S. J. McGowan[1], S. J. L. Knight[1], S. M. A. Swagemakers[3], D. J. Downes[1], R. Schwessinger[1], J. R. Hughes[1], P. Hublitz[1], J. A. Sloane-Stanley[1], D. Korona[1], J. A. C. Goos[3], M. F. van Dooren[3], D. Johnson[4], S. A. Wall[4], J. E. V. Morton[5], L. C. Wilson[6], K. E. Heath[7], J. C. Taylor[1], I. M. J. Mathijssen[3], S. R. F. Twigg[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Genomics plc, Oxford, United Kingdom, [3]Erasmus Univ. Med. Ctr., Rotterdam, Netherlands, [4]Oxford Univ. Hosp. NHS Fndn. Trust, Oxford, United Kingdom, [5]Birmingham Women's Hosp. NHS Fndn. Trust, Birmingham, United Kingdom, [6]Great Ormond Street Hosp. NHS Fndn. Trust, London, United Kingdom, [7]INGEMM & UMDE; IdiPAZ, Hosp Univ La Paz & CIBERER, Madrid, Spain

Abstract:

Background Identification of novel monogenic causes of craniosynostosis (premature fusion of cranial sutures) has greatly accelerated through the implementation of next-generation sequencing, but outside the coding regions, detecting and interpreting the pathogenic implications of sequence changes is challenging. The organisation of the 3D genome into regulatory units termed “topologically-associating domains” (TADs), has helped to explain the molecular pathology underlying clinically significant chromosomal rearrangements. Here, we aimed to identify the underlying cause in genetically undiagnosed craniosynostosis cases, including two of the largest unsolved multi-generation craniosynostosis families in the UK. **Methods** We performed whole-genome sequencing (WGS) of index cases, capture-based targeted resequencing of >320 craniosynostosis patients with unknown etiology, capture-C using patient cells and single-cell transcriptomics using embryonic craniofacial tissue from a CRISPR/Cas9-engineered mutant mouse model. **Results** Using WGS we identified heterozygous rearrangements at chr1p31.3 that fully co-segregate with craniosynostosis in 11 affected individuals from 3 independent families, and we characterised two further rearrangements by targeted resequencing. These lesions, an 11.5 kb tandem duplication and four overlapping deletions (≥275 kb) flank *FOXD3*, encoding a pioneer winged-helix transcription factor critical for early development of embryonic stem cells and neural crest, but not previously implicated in craniosynostosis. The deletions all lie downstream of *FOXD3* and remove a TAD boundary; both computational predictions and capture-C experiments support the fusion of adjacent TADs, potentially leading to *FOXD3* mis-expression. The duplication lies in non-coding sequence ~193 kb upstream of *FOXD3;* this encompasses a highly conserved 1.3 kb region previously shown to act as a *Foxd3* enhancer in chicken neural crest. We generated a mouse containing the equivalent duplication ("DUP"), which manifests synostosis of the lambdoid and coronal sutures in 84% and 20% of homozygotes, respectively, supporting the pathogenicity of the duplication in humans. Single-cell transcriptomics data from dissected anterior fontanelles of E17.5 DUP embryos shows striking differences in the mutant animals compared to wild-type mice, including apparent ectopic cells with a neuronal signature. **Conclusion** Our findings uncover a new disease locus involved in craniosynostosis and demonstrate a mechanism whereby non-coding copy number variations lead to this phenotype by mis-expression following neural crest determination.

S64. New molecular and analytical tools to study Mendelian disorders

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 412. MARRVEL-AI: Knowledge based artificial intelligence for variant pathogenicity prediction for Mendelian disorders

Authors:

D. Mao, C. Liu, L. Wang, R. AI-Ouran, L. Li, S. Kim, S. Pasupuleti, C. Deisseroth, P. Liu, B. Yuan, Members of UDN, S. Yamamoto, M. Wangler, B. Lee, F. Xia, L. Meng, Baylor Genetics, H. Bellen, Z. Liu; Baylor Coll. of Med., Houston, TX

Abstract:

Abstract:Every year thousands of patients, with potential rare genetic disorders, face uncertainty when healthcare providers are unable to discover the cause for their symptoms. The Undiagnosed Diseases Network (UDN) seeks to provide answers for patients and families affected by these mysterious conditions. The process of defining pathogenicity currently requires labor-intensive manual searches of a variety of databases and web resources. This manual process is time-consuming, subject to inter-user variability and variations in the depth or quality of the databases. It also requires broad expertise across multiple biological and informatics domains. Here, we created a systematic, comprehensive search engine, MARRVEL (Model organism Aggregated Resources for Rare Variant ExpLoration, http://marrvel.org), that mines all the critical information for variant analysis and presents it in a succinct, user-friendly way. MARRVEL integrates human databases (OMIM, gnomAD, ExAC, ClinVar, Geno2MP, DGV, and DECIPHER) and seven model organism databases from yeast to mammals. Furthermore, we are also developing a Knowledge-based Artificial Intelligent system (MARRVEL-AI) to prioritize and identify novel disease-causing coding variants. The interpretability of a machine learning method inversely correlates with its accuracy for complex tasks. To circumvent this, we are combining different models of artificial intelligence with complementary strengths, such as expert system and random forest. With only a small training data set, our model out-perform all compared methods on validation date set. In addition, our MARRVEL-AI also provide confidence score at case level where users can preselect cases for review.

S64. New molecular and analytical tools to study Mendelian disorders

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 413. Determining the parent of origin of homologous chromosomes without parental data

Authors:

V. Hanlon[1], V. Akbari[2,3], K. O'Neill[2], L. Lefebvre[3], K. Schrader[3,4], P. Lansdorp[3,1], S. Jones[2,3]; [1]Terry Fox Lab., BC Cancer Agency, Vancouver, BC, Canada, [2]Canada's Michael Smith Genome Sci. Ctr., BC Cancer Agency, Vancouver, BC, Canada, [3]Dept. of Med. Genetics, Life Sci. Inst., Univ. of British Columbia, Vancouver, BC, Canada, [4]Dept. of Molecular Oncology, BC Cancer, Vancouver, BC, Canada

Abstract:

Background: Single-cell template strand sequencing (Strand-seq) has been shown to provide sparse chromosome-length haplotype information that, in combination with long-read sequence data, can be used for accurate genome-wide haplotyping. However, neither Strand-seq nor other state-of-the-art phasing approaches can infer whether alleles were inherited from the biological mother or the father. Such inferences are of interest because they could be used as a shortcut for cascade genetic testing that traces pathogenic variants through families. Imprinted differentially methylated regions—known autosomal loci where alleles are methylated only if they are inherited from the mother (or vice versa, from the father)—are one potential source of parent-of-origin information. Such imprinting is generally consistent in different tissues, individuals, and human populations.
Aim: To determine whether combined nanopore and Strand-seq data can be used to infer the parent of origin for genetic variation without parental sequencing.
Method: We combined DNA methylation and DNA sequence from long nanopore reads with Strand-seq data from publicly available samples and datasets. Strand-seq was used to phase nanopore reads, which were used to phase nanopore-detected DNA methylation, SNVs, and indels into two haplotypes per autosome. Each haplotype was assigned a parent of origin by examining phased DNA methylation at an assembled catalogue of imprinted differentially methylated regions.
Results: Parent of origin was correctly inferred for all autosomes with an average mismatch error rate of 0.31% for SNVs and 1.89% for indels.
Conclusion: We describe and validate a method that combines nanopore long read sequence data with phasing using Strand-seq data to distinguish maternal and paternal alleles without parental data with 99% accuracy.

S64. New molecular and analytical tools to study Mendelian disorders

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 414. The use of optical genome mapping in genetically unsolved neurodevelopmental disorders

Authors:

I. Schrauwen[1], A. Acharya[1], N. S. Lin[1], S. M. Leal[1], H. Kokkonen[2], I. Järvelä[3]; [1]Columbia Univ., New York, NY, [2]Oulu Univ. Hosp. and Univ. of Oulu, Oulu, Finland, [3]Univ. of Helsinki, Helsinki, Finland

Abstract:

Optical genome mapping (OGM) is an innovative technique with the capability to identify structural variants (SVs) that are undetectable or difficult to detect via short-read sequencing techniques. It fluorescently tags long linearized DNA molecules at specific sites to create a detailed map of genomic variation including repetitive regions. This study aims to investigate the use of OGM in the molecular diagnosis of neurodevelopmental disorders (NDDs). We selected families previously investigated via conventional exome sequencing techniques with no conclusive causative variant(s) identified. All affected individuals have a severe and disabling NDD with intellectual disability and other abnormal features of neurodevelopment. So far, we have completed OGM in 30 families with a genetically unsolved NDD and analysis of additional families is ongoing. The results of the larger cohort will be presented. Ultra-high molecular weight DNA was extracted from frozen blood, fluorescently labeled with DLE-1 and OGM was performed on Bionano's Saphyr instrument. Single-molecule maps were assembled *de novo* into consensus maps using the Bionano Solve data analysis software, followed by SV calling against the hg38/grch38 reference. The average filtered molecule N50 was 279kbp and the average effective coverage of the reference was 189x. In 6 families (20%), we found possible pathogenic SVs in known NDD genes, including *OPHN1*, *KCNQ3, NUP133, H3F3A, SON,* and *PHF8*. These variants are currently being validated and reconstructed further with techniques such as targeted long-read sequencing. In 4 families, we found rare or unique SVs in novel candidate genes. Variants in known NDD genes or candidate variants of interest missed by exome sequencing mainly consisted of larger insertions (>1kbp), inversions, and deletions/duplications of a low number of exons (1-4 exons). In conclusion, this study aims to take a more comprehensive view into the genomic landscape of rare variants implicated in NDDs via OGM. In addition to improving molecular diagnosis, this technique may also reveal novel NDD genes which may harbor complex SVs often missed by standard sequencing techniques.

S64. New molecular and analytical tools to study Mendelian disorders

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 415. Reducing uncertainty and improving diagnostic yield in the MSSNG cohort using DNA methylation signatures

Authors:

S. Choufani, M. Jangjoo, M. Reuter, S. Scherer, R. Weksberg; Genetics and Genome Biology Program, Res. Inst., The Hosp. for Sick Children, Toronto, ON, Canada

Abstract:

Genome sequencing captures many types of inter-individual genomic differences; however, our ability to identify variants that are disease-causing can be challenging. In many cases, genetic variants having uncertain significance (VUS) are reported. Although published guidelines exist to interpret VUS, diagnostic and research laboratories often rely on *in silico* prediction tools to evaluate the pathogenicity of VUS. Several studies have demonstrated that the specificity of these tools can be as low as 13% with poor correlation across results from different tools. Our recent research has shown that DNA methylation (DNAm) signatures have diagnostic utility with high specificity and sensitivity to classify VUS in genes that function in epigenetic regulation. The goal of this study was to test the utility of DNAm signatures in classifying VUS in the Autism Spectrum Disorder (ASD) MSSNG cohort, a massive open-source whole genome sequencing project. Here we present our findings on utilizing DNAm signatures for 21 specific genes/disorders to reduce uncertainty and improve diagnostic yield in the MSSNG cohort. This study was approved by the Research Ethics Board (REB) at the Hospital for Sick Children. A review of MSSNG sequencing data for over 10,000 individuals identified ~70 individuals with VUS in 21 genes for which we have validated DNAm signatures. DNAm profiles for these ~70 individuals were generated using the Illumina Infinium MethylationEPIC BeadChip (~850,000 CpG sites genome-wide). Variants were classified using gene-specific machine learning models as part of the analytical pipeline established using gene specific disease (n=136) and control references (n=125). We classified seven variants in *CHD8*, *CHD7*, *DNMT3A* and *DNMT1* as pathogenic, 49 variants as benign, and four variants as intermediate. Intermediate classification which can reflect mosaicism, atypical clinical presentation or a pathogenic variant in a related gene can be valuable in defining downstream workflow. Currently, detailed clinical review is underway to assess if the DNAm classification is consistent with each individual's clinical presentation and to further investigate variants with intermediate classification. Our results show that functional DNAm signatures can enhance the diagnostic yield for genomic disorders.

S65. Population screening: From patient identification to return of results

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 418. Assessing a legal pathway to implement a cascade traceback screening program for ovarian cancer

Authors:

J. Wagner[1], J. Tanniru[1], C. Chane[1], M. Meyer[2]; [1]Penn State Univ., University Park, PA, [2]Geisinger, Danville, PA

Abstract:

The National Cancer Institute recently funded projects to explore a "traceback" strategy for providing at-risk relatives of individuals who previously were diagnosed with ovarian cancer an opportunity to learn of their genetic risk information through cascade screening. Applicable laws can facilitate or frustrate these efforts to communicate genomic risk information to relatives. ELSI scholars have examined the HIPAA Privacy Rule and its implications for patient- and provider-mediated familial risk notification and cascade genetic testing; however, whether the HIPAA Privacy Rule's Public Health Exception (PHE) offered a lawful pathway for direct contact between providers and at-risk relatives remained unclear. We explored the contours of the PHE to examine whether it offers a viable pathway for implementation of a traceback program for ovarian cancer. If viable, such a pathway would help providers overcome challenges when a patient-pro band is unable or unwilling to give consent or authorization for disclosure of the information. We explored the text of the HIPAA PHE to determine the legal elements that must be met for the exception to be applicable. We then examined reportable conditions among the 50 states, by manually reviewing the state regulations and by consulting the State Reportable Conditions Assessment (SRCA) database, to determine whether cancer is recognized as a reportable condition. Finally we performed a 50-state survey of the genetic information privacy restrictions to determine whether, even if cancer is a reportable condition, direct contact between healthcare providers and at-risk individuals is forbidden. Manual review of state regulations showed cancer on the reportable conditions lists of eight states, and the SRCA database showed 31 jurisdictions as having explicit cancer reporting requirements-a discrepancy presumably due to distinctions between general reportable disease regulations and distinct cancer registry regulations. Even where ovarian cancer is a reportable condition, we identified other barriers that preclude direct contact of at-risk relatives by healthcare providers who are not employed by a public health authority. We conclude the PHE is not a viable pathway for implementation of an ovarian cancer cascade traceback program.

S65. Population screening: From patient identification to return of results

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 419. The Precision Population Health Initiative, scaling genomic medicine into population health

Authors:

R. Green[1,2,3], N. Encina[4,2], C. L. Blout Zawatsky[1,2], I. Miller[5,2], E. Drake[1,2], S. Adelson[1,2], L. Subramanian[5,2], J. Zenger-Hain[6], L. Onelio[7,2], A. Bitton[5,2,3,7]; [1]Brigham and Women's Hosp./Genomes2People, Boston, MA, [2]Ariadne Labs, Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]Harvard Sch. of Publ.h Hlth., Boston, MA, [5]Brigham and Women's Hosp., Boston, MA, [6]Beaumont Hlth., Dearborn, MI, [7]Harvard Sch. of Publ. Hlth., Boston, MA

Abstract:

Purpose: Information gained from genomic testing has begun to transform patient care in specialty fields like oncology and cardiology, allowing for tailored medical treatment and management. As technology improves, and costs decline, proactive genomic screening (PGS) has become a reality. PGS has the ability to identify individuals at risk of developing disease before symptoms present, allowing for surveillance, earlier detection, and better prevention or treatment. However, for PGS to achieve its full potential at scale, it must diffuse into primary care. The Precision Population Health Initiative (PPH) aims to address this by uniting early studies in implementation science with technologies designed to implement PGS at scale. Methods: The PPH team developed a PGS assessment process (including surveys and provider interviews) to assess health systems' readiness for successful implementation of genomic medicine. Assessment learnings are used to create PGS implementation toolkits, which include guidance on genomic test selection, patient and provider educational materials, intuitive models for implementing new clinical pathways, and scalable tools to make genomic medicine more accessible at the point of care. Toolkit development progresses through stages of design, test, and spread, and is shaped by an interdisciplinary team of experts in implementation, genomics, clinical care, innovation and communication, as well as by the perspectives of providers and patients. In 2021, we piloted the PGS readiness assessment process with Beaumont Health, surveying 330 physicians across 50 practices and 35 executive leaders across 11 system-level groups, and conducting qualitative interviews with 15 specialists. Results: Our pilot showed that this readiness assessment process can provide valuable insights for PGS implementation planning and toolkit development. The assessment found that most providers and leaders saw the value of a PGS program (88% of leaders were passionate about this work, 78% of providers thought PGS would benefit their patients) and highlighted resources and strengths that will facilitate program implementation. It also helped flag key areas that required preparatory attention, such as: shortage of resources and support for physicians (guidelines, consent processes, sample collection/coordination, post-test referral pathways, etc); physician and leader buy-in; genomics staffing and support; and patient-related concerns, such as cost and privacy. Conclusion: The PPH PGS readiness assessment process can successfully be used to inform design of a system-wide PGS implementation roadmap and toolkit.

S65. Population screening: From patient identification to return of results

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 420. Geno4ME: Establishment of an equitable whole-genome sequencing-based platform for clinical screening in a large healthcare system

Authors:

J. T. Wagner[1], I. A. Lucas Beckett[2,1], K. R. Emery[1], B. A. Cosgrove[1], J. T. Welle[1], K. Jade[2], N. D. Wagner[1], J. Rinaldi[1], L. Dickey[1], K. G. Jones[1], L. Yuan[1], E. M. Shull[1], J. Clemens[1], K. Vartanian[1], A. T. Magis[2], M. B. Campbell[1], B. D. Piening[1,3], C. B. Bifulco[1], O. K. Gordon[1,4]; [1]Providence Hlth.System, Portland, OR, [2]Inst. for Systems Biology, Seattle, WA, [3]Earle A. Chiles Res. Inst., Portland, OR, [4]St John Cancer Inst., Santa Monica, CA

Abstract:

Recent population-wide genetic testing analyses suggest that over 1% of the general population are carriers for a clinically actionable single-gene genetic condition. In addition, approximately 80% of variability in drug efficacy and safety are attributed to an individual's pharmacogenomics (PGx) profile. Here, we report the establishment of the Genomic Medicine for Everyone (Geno4ME) program across the highly diverse seven-state Providence Health System. Key components include targeted and multi-lingual outreach to traditionally underrepresented groups, a novel e-consent and education platform, and whole genome sequencing with clinical return of results via the Electronic Health Record (EHR) for 78 hereditary disease genes and four clinically relevant pharmacogenomics (PGx) genes. The program provides genetic counseling and pharmacist support for patients and educational resources and peer-to-peer support to assist providers in caring for patients with a positive result. Over the initial months of the study, over 23,000 potential participants were outreached; of this, 1,971 were consented to the study (of which 48.4% are people of color: 15.4% Hispanic, 14.9% Asian, 7.9% more than one race, 7.2% Black, 3.0% other) and 753 have had results returned so far. Fifty-two (7.0%) initial participants were found to have an actionable gene variant in the hereditary disease panel. Additionally, 111 (14.7%) of initial participants were currently taking one of the supported medications with a medical recommendation resulting from PGx genotype. Overall, 20.5% of initial participants had a test result with one or more medical intervention recommendations. Geno4ME plans to enroll up to 5,000 total participants within the next year and plans to expand the proportion of the genome that is included in the clinical deliverable and scheduled re-evaluation of variants for clinical significance. We propose this model as a viable, effective framework for population screening into routine healthcare and the whole-genome sequencing platform as key for continued reanalysis and long-term integration into routine clinical practice. This model specifically addresses issues of recruitment of diverse populations in genomics research and incorporating primary care provider education and genomics fluency into core program delivery.

S65. Population screening: From patient identification to return of results

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 421. Genome-first evaluation with exome and clinical data uncovers underdiagnosis of genetic disorders in a large healthcare system

Authors:

I. Forrest, A. Duffy, J. Park, H. Vy, L. Pasquale, G. Nadkarni, J. Cho, R. Do; Icahn Sch. of Med. at Mount Sinai, New York, NY

Abstract:

Population-based genomic screening may help diagnose individuals with variants in disease-predisposition genes. However, the utility of a genome-first approach to identify clinically undiagnosed individuals with evidence of disease for most genetic disorders is unclear. Here, we performed a genome-first evaluation for 10 genetic disorders in a clinical care cohort of 29,039 participants with linked exome and electronic health record (EHR) data. A total of 1,871 individuals with 680 pathogenic/likely pathogenic or loss-of-function (P/LP/LoF) variants were identified, yielding 1,965 variant-disease observations; 1,643 (84%) lacked a corresponding clinical diagnosis in the EHR. A subset of 153 (9.3%) observations had EHR evidence of symptomatic disease, but were not clinically diagnosed or medicated. These were identified for nine out of 10 disorders (90%), including familial hypercholesterolemia (3/8 [38%] undiagnosed observations with evidence of disease), prostate cancer (38/138 [28%]), and cardiomyopathy (68/453 [15%]). We determined the proportion of individuals with P/LP/LoF variants in each gene who had disease (i.e., penetrance) and observed a greater proportion of individuals without a clinical diagnosis who had evidence of disease for those with variants in highly penetrant genes compared to those with variants in weakly penetrant genes (19% versus 3.6% for genes with >50% versus 0% observed penetrance, respectively). As a demonstration of the clinical utility of population-based genomic screening, we conducted a targeted phenotype evaluation of individuals with variants in highly penetrant genes to reveal new disease diagnoses. The finding of P/LoF *PKP2* variants informed an assessment of cardiac phenotypes that supported new diagnoses of cardiomyopathy, while the identification of *LDLR* P/LP/LoF variants guided an examination of cardiovascular and lipidic profiles that uncovered new diagnoses of familial hypercholesterolemia. The prevalence of P/LP/LoF variants exceeded that of clinical diagnoses in a large biobank population and a subset of clinically undiagnosed individuals were identified with evidence of disease once genetic information was used. These results demonstrate the potential of population-based genomic screening in healthcare systems.

S65. Population screening: From patient identification to return of results

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 422. Tests and procedure rates following return of medically actionable monogenic variants within the eMERGE cohort

Authors:

J. E. Linder[1], R. Tao[1], I. J. Kullo[2], D. L. Veenstra[3], R. Sharp[2], C. J. Lee[2], W. K. Chung[4], L. J. Rasmussen-Torvik[5], R. Rowley[6], K. A. Leppig[7], C. Hoell[5], C. A. Prows[8], N. Wilmayani[1], M. S. Williams[9], eMERGE Network, J. F. Peterson[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Mayo Clinic, Rochester, MN, [3]Univ. of Washington, Seattle, WA, [4]Columbia Univ., New York, NY, [5]Northwestern Univ., Chicago, IL, [6]Natl. Human Genome Res. Inst., Bethesda, MD, [7]Kaiser Permanente of Washington, Seattle, WA, [8]Children's Hosp. Med. Ctr., Cincinnati, OH, [9]Geisinger, Holmen, WI

Abstract:

The eMERGE Network sequenced 25,000 participants for actionable monogenic findings; results were returned to participants and providers, and placed into electronic health records (EHRs) of 10 health systems. We sought to accurately capture changes in adult participants' health care utilization, including screening tests, diagnostic tests, and diagnostic procedures, after the return of sequencing results (RoR). We developed a knowledgebase of Current Procedural Knowledge (CPT®) codes to track health services in a high-throughput manner delivered for five conditions associated with monogenic variants: arrhythmia (arrhyth), breast cancer (bc), cardiomyopathy (cardio), colorectal cancer (crc), and familial hypercholesterolemia (fh). CPT codes were manually grouped into similar test or procedure types and validated using manual chart review. We analyzed the rate of tests or procedures per year for patients with pathogenic or likely pathogenic (P/LP) variants and those without. Participants were 1:1 matched for age, sex, and enrollment site and stratified by condition. We compared pre- and post-RoR test rates among individuals with P/LP findings; secondly, we conducted a difference in difference analysis incorporating pre- and post- test rates among matched participants without a variant to account for secular effects. Overall, 321 unique CPT codes were categorized into 73 code clusters each representing a single type of test or procedure (# clusters/condition): arrhyth (18); bc (18), cardio (26), crc (8), fh (3). 479 participants received P/LP findings (N = participants with P/LP): arrhyth (95), bc (96), cardio (96), crc (105), and fh (87) and were matched to 472 participants without variants. When analyzing rates of services (tests and procedures) in P/LP participants pre- and post-RoR, there were significantly more services delivered post-RoR (median rate of services per year of observation [IQR]; p-value): arrhyth (pre: 0.9 IQR [0, 2.55] vs post: 2.5 IQR [0, 5.43], $p < 0.0001$), bc (pre: 0 [0, 0.86] vs post: 0 [0, 1.95]; $p = 0.01$), cardio (pre: 1.8 [0, 4.3] vs post: 3.5 (0.76, 7.4); $p = 0.001$), and crc (pre: 0 [0, 0] vs post: 0 [0, 1.2]; $p = 0.007$), fh was not significantly different. A difference in difference analysis examining the change in services among participants with P/LP results to those without a variant found significantly higher rates in the P/LP group for three (arryth, cardio, crc) of the five conditions. Using a high-throughput approach for large scale outcomes assessments in de-identified datasets, we demonstrated that return of medically actionable variants significantly increased utilization of health services related to return of results.

S65. Population screening: From patient identification to return of results

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 423. Addition of chatbot to the return of genetic results process for biobank participants

Authors:

N. Larson[1], K. Kolbert[1], B. C. Thomas[1], J. A. Anderson[1], J. L. Kemppainen[1], E. Simmons[2], E. Ryu[1], R. Gupta[1], J. T. Bublitz[1], P. R. Chinthalapelly[1], Z. Fredericksen[1], G. D. Jenkins[1], C. R. Kirt[1], T. Kruisselbrink[1], J. Cerhan[1], S. N. Thibodeau[1], J. E. Olson[1]; [1]Mayo Clinic, Rochester, MN, [2]Invitae, San Francisco, CA

Abstract:

Introduction Large-scale return of genomic results to participants is challenging both in terms of genetic counselor time and the infrastructure surrounding them. The Mayo Clinic Biobank recently completed exome sequencing on 54,000+ participants. The Biobank policy is to return pathogenic/likely pathogenic (P/LP) variants in the 73 genes listed by the American College of Medical Genetics (ACMG73). Based off mutation prevalence in healthy populations, we anticipate needing to offer results to 2000-4000 participants. To increase the capacity of our genetic counseling (GC) team and prepare for the influx of results needing return, we conducted a pilot study that used a chatbot system to schedule and prepare subjects for their GC visit. We report our initial experiences with the chatbot in this pilot study, evaluating the rate of usage, along with feasibility and satisfaction among our participants. **Methods** For the pilot, pathogenic and likely pathogenic variants in selected genes were identified in 91 subjects from two prior small studies (N=3482). Participants are sent invitations via mail followed by email and text to utilize the chatbot to obtain basic education on genetics, provide their personal and family health history, and schedule a genetic counseling appointment. To assess the participant's experiences using the chatbot to schedule their appointment, participants were prompted to answer a short survey within the chatbot. Participants were also able to opt out of the chatbot and access traditional methods of scheduling and education. Frequencies and counts were calculated using Excel. **Results** The pilot study is currently ongoing. In the first round of results, 34 invitation letters were sent and 32 (94%) opted to receive their results. One participant declined due to health reasons and another did not respond after multiple contacts. Of the 32 responders, 5 scheduled a GC visit via phone and the remaining 27 (87.1%) completed education, family and personal health history entry, and appointment scheduling through the chatbot. Among chatbot users, 96% reported being "somewhat satisfied" or "very satisfied" with the process and reported finding scheduling via the chatbot to be "very easy". The age range for the chatbot users was 42-83, with a mean of 64.2. The ages for those that scheduled over the phone were 77-90 with a mean of 83.2. **Conclusions** Based upon preliminary data, patient satisfaction and uptake of the chatbot was very high even among older patients. The use of chatbots and other methods of pre-appointment communication will be vital for the feasibility of large-scale return of results projects without limiting biobank participant satisfaction.

S66. Towards the defeat of neurodegenerative disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 426. Polygenic architecture of a novel MRI endophenotype: “The hemochromatosis brain”

Authors:

R. Loughnan[1], J. Ahern[1], M. Boyle[1], T. Jernigan[1], A. Dale[1], W. Thompson[1,2], A. Schork[3], C. Chieh Fan[1,2]; [1]Univ. of California, San Diego, La Jolla, CA, [2]Laureate Inst. for Brain Res., Tulsa, OK, [3]Res. Inst. for Biological Psychiatry, Roskilde, Denmark

Abstract:

Iron is an important element required for the body to function properly; however, an excess in healthy tissue can lead to cell damage. C282Y is a genetic variant prevalent in European populations which is known to increase iron levels in the body and is responsible for the majority of hemochromatosis cases. Recent work from our group analyzing brain Magnetic Resonance (MR) images established, in a large sample of older adults, that individuals who were homozygote for C282Y showed evidence consistent with substantial iron deposition localized to brain regions related to movement. These individuals also had nearly two-fold increased risk for developing movement disorders, like Parkinson’s Disease. Here, we describe the broader polygenic architecture of this novel endophenotype - the “hemochromatosis brain.” This multivariate MRI phenotype captures average brain-wide differences of C282Y homozygote individuals and may reflect consequences of iron deposition localized to motor circuits. To do this we trained a classifier in a sample of 836 individuals from the UKBiobank to distinguish C282Y homozygotes from other genotypes using phenotypes derived from MR images sensitive to iron. This classifier was then applied to an independent set of 25,716 individuals for which MRI were available in the UKBiobank. This generated, for each individual, a PolyVoxel Score (PVS) that captures variation in MRI phenotypes associated previously with C282Y homozygosity and may reflect the degree of brain-specific iron deposition or other disease mechanisms. We show this “hemochromatosis brain” PVS to be highly heritable (h_{snp}=0.41, SE=0.03). Performing GWAS in these 25,716 individuals identified 45 genome wide significant loci. Many of these signals are proximal to genes known to be involved in iron homeostasis (TF, rs4428180, p=$1.88x10^{-56}$; HFE rs144861591, p=$6.37x10^{-51}$; TMPRSS6, rs2072860, p=$8.57x10^{-45}$). In addition, we found signals near genes previously implicated in uric acid levels (SLC17A3, rs62394273, p=$1.38x10^{-14}$; ABCDG2, rs1481012, p=$2.90x10^{-9}$) which is of interest as uric acid is thought to play an antioxidant role in reducing risk of Parkinson’s disease. Furthermore, the PVS showed a trending association with increased risk for Parkinson’s disease (OR=1.3, p=0.057). In this study we show that the multivariate MRI phenotype associated with C282Y homozygosity is, in fact, highly heritable, polygenic, and may represent a novel brain endophenotype of Parkinson’s disease. Our approach of combining classical forward and reverse genetics approaches to define and describe potentially disease mediating endophenotypes is unique and could be applied more broadly.

S66. Towards the defeat of neurodegenerative disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 427. Immune system and blood-brain barrier-wide biomarker analyses provide causal evidence for autoimmunity in dementia causing diseases

Authors:

J. Lindbohm[1], N. Mars[2], P. Sipilä[1], A. Singh-Manoux[3], H. Runz[4], FinnGen, G. Livingston[3], S. Seshadri[5], R. Xavier[6], A. Hingorani[3], S. Ripatti[2], M. Kivimäki[3]; [1]Univ. of Helsinki, Helsinki, Finland, [2]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, [3]Univ. Coll. London, London, United Kingdom, [4]Univ Heidelberg, Heidelberg, Germany, [5]Boston Univ., Boston, MA, [6]Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract:

Introduction
Immune system and blood brain barrier (BBB) dysfunction are implicated in the development of Alzheimer's and other dementia-causing diseases, but their causal role remains unknown. We used triangulation to identify novel disease etiology, drug targets, and medications with potential for repurposing in dementia-causing diseases by conducting six complementary studies.
Methods and Results
The first study performed a large-scale analysis for 43,643 immune system and BBB-related biomarkers using two-step Mendelian randomization and identified 126 potential causal risk factors for dementia-causing diseases. The second study used pathway analyses and linked the biomarkers to amyloid-β, tau and α-synuclein pathways and to autoimmunity-related processes. Based on these data we hypothesized that autoimmunity plays a role in diseases causing dementias. We then used studies three to six to triangulate consistency of this hypothesis using methods with different sources of bias.
The third study constructed MR-based polygenic risk score (MR-PRS) based on the SNPs that associated with the 126 biomarkers in MR analyses. We then conducted a phenome-wide analysis using the MR-PRS in FinnGen study (N=339,233) which identified strong common genetic background for dementias and autoimmune diseases (p-value < 10^-75 for dementias, type 1 diabetes and rheumatoid arthritis). These analyses were supported by the fourth study that used HLA-allele-wide analyses to identify nine HLA-risk alleles for dementia-causing diseases in the FinnGen study (N=339,233). In the fifth study, we used inverse-probability-weighted analyses in the FinnGen study (N=90,512) to simulate randomized controlled drug trials in observational data, and showed that anti-inflammatory methotrexate reduced Alzheimer's and Parkinson's disease risk (HR=0.64, 95%CI 0.49-0.88 and 0.20, 0.05-0.83) in those with higher MR-PRS for the biomarkers. The sixth study using plasma proteins in the Whitehall II study (N=6,235) showed that plasma IFIT2 may be a drug target for dementia causing diseases independent of APOE gene status.
Conclusions
These converging results from six different research lines suggest that autoimmunity is a modifiable component in diseases causing dementias. Using immune-system-dysfunction related MR-PRS we identified a novel subgroup of individuals whose dementia risk may be reduced with anti-inflammatory medications.

S66. Towards the defeat of neurodegenerative disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 428. Genetic analysis of multiple sclerosis severity identifies a novel locus and implicates CNS resilience as a major determinant of outcome

Authors:

A. Harroud[1], International Multiple Sclerosis Genetics Consortium, MultipleMS, S. J. Sawcer[2], S. E. Baranzini[1]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Univ. of Cambridge, Cambridge, United Kingdom

Abstract:

Multiple sclerosis (MS) is an autoimmune disease of the central nervous system (CNS) that results in significant neurodegeneration in the majority of those affected and is a common cause of chronic neurological disability in young adults. Despite successful genetic characterization of MS susceptibility, genetic associations with severity and prognosis have yet to be identified. To provide insight into the mechanisms driving disease progression, we conducted a genome-wide association study of the age-related MS severity score in 12,584 cases with long-term outcomes (mean disease duration 18.2 years) and replicated our findings in a further 9,805 cases. We identified a significant association with rs10191329 in the *DYSF-ZNF638* locus (P=3.6×10^{-9}), the risk allele shortening the median time to require a walking aid by 3.7 years. Carriers displayed faster disability accrual across 54,113 study visits. The lead variant also associated with faster 24-week confirmed disability worsening, a metric used as the primary outcome in MS progression therapeutic trials. We identified an additional suggestive association with the low-frequency variant rs149097173 in the *DNM3-PIGC* locus (P=2.3×10^{-7}), with a similarly consistent effect on longitudinal MS-specific outcomes in survival and linear mixed model analyses. Statistical fine-mapping identified putative causal variants at these loci. Prioritized genes displayed shared cell-type specificity for oligodendrocyte lineage cells. These cells play a key role in MS pathophysiology, including axonal remyelination after injury. We estimated SNP-heritability at 13%, explaining some of the considerable variability in MS outcome. We also identified significant enrichment for gene expression in brain and spinal cord tissues, in marked contrast to the pronounced immune signal seen for MS susceptibility. Furthermore, the genetic architectures of MS susceptibility and severity were found to be largely distinct by genetic correlation and polygenic risk score approaches. Last, Mendelian randomization analyses indicated a protective role for higher educational attainment (P=9.7×10^{-4}), consistent with findings in neurodegenerative conditions such as Alzheimer's disease. In summary, we identified the first genetic modifier of MS severity, offering novel potential therapeutic targets with genetic support. In contrast to immune-driven susceptibility, these findings indicate a key role of CNS resilience and neurocognitive reserve in determining outcome in MS.

S66. Towards the defeat of neurodegenerative disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 429. Single-cell multi-region dissection of Alzheimer's disease progression reveals selectively vulnerable neurons

Authors:

C. Boix[1], H. Mathys[1], L. Akay[1], A. Ng[2], X. Jiang[3], K. Galani[1], J. Mantero[1], N. Band[4], B. James[1], J. Davila-Velderrain[1], D. Bennett[5], L-H. Tsai[1], M. Kellis[1]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Univ. of California, Los Angeles, Los Angeles, CA, [3]Univ. of California, San Francisco, San Francisco, CA, [4]Univ. of Oxford, Oxford, United Kingdom, [5]Rush Univ. Med. Ctr., Chicago, IL

Abstract:

Alzheimer’s disease (AD) is the primary cause of dementia worldwide, but its molecular mechanisms across genes, pathways, cell types, and brain regions remain poorly understood. To address this, we generated a multi-region single-cell transcriptomic atlas of the aged human brain covering 1.3 million cells from 283 post-mortem human brain samples across six brain regions, 48 individuals, and six stages of AD progression. We identify 76 cell types, with region-specific subtypes of excitatory and inhibitory neurons and glia, including an inhibitory interneuron population unique to the thalamus and distinct from canonical parvalbumin, somatostatin, and VIP inhibitory subclasses. We discover thalamic astrocyte subtypes absent from the neocortex and characterized by specific neurotransmitter transporters, ion channels, and signaling receptor expression. We develop a novel multi-resolution method to discover gene expression modules, revealing region-specific glial programs. We report glial compositional and transcriptomic differences tied to specific regional pathology burden, stage of AD progression, or cognitive status. We identify region-specific depleted excitatory subtypes, gene expression programs that may mediate their resilience and vulnerability in AD, and pinpoint vulnerable connected components of the brain's neuronal circuitry in persons with AD. Our work charts a spatial map of AD and brain architecture that provides insight into cellular response and resilience to disease progression and offers a valuable reference for future studies of the human brain.

S66. Towards the defeat of neurodegenerative disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 430. Regional genetic correlations highlight relationships between neurodegenerative diseases and the immune system

Authors:

F. Lona Durazo[1], R. Reynolds[2], S. Scholz[3], M. Ryten[4], S. Gagliano Taliun[1]; [1]Montréal Heart Inst., Montreal, QC, Canada, [2]Univ. Coll. London, London, United Kingdom, [3]NIH, Bethesda, MD, [4]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom

Abstract:

Neurodegenerative diseases (ND), including Alzheimer's disease (AD), Parkinson's disease (PD), Lewy body dementia (LBD) and amyotrophic lateral sclerosis (ALS), are devastating complex diseases that result in a physical and psychological burden to patients and their families. There have been significant efforts to understand the genetic basis of ND resulting in the identification of disease risk-associated variants involved in several molecular mechanisms, including those that influence immune-related pathways. Regional genetic correlations, in contrast to genome-wide correlations, among immune-neurodegenerative trait pairs have not been comprehensively explored, but could shed light on additional immune-mediated risk-associated loci. Here, we systematically assessed the potential role of the immune system in neurodegeneration, by estimating regional genetic correlations between ND and immune-cell-derived single-cell expression quantitative trait loci (sc-eQTLs), using the recently developed method of Local Analysis of [co]Variant association (LAVA). We used the most recently published GWAS for six ND and publicly available sc-eQTLs derived from 120 individuals from the sc-eQTLGen Consortium, capturing aspects of the innate and adaptive immune systems. Additionally, we tested GWAS from well-established autoimmune diseases, Crohn's disease (CD) and ulcerative colitis (UC), and the immune-mediated neurodegenerative disease, multiple sclerosis (MS), as positive controls. In addition to showing significant global genetic correlations (Bonferroni corrected p-value < 0.05) across our positive control GWAS, we observed significant correlations between PD and LBD (r_g = 0.65; p-value = 0.001). Regional genetic correlations highlighted positive local genetic correlations across ND, including loci containing the genes *BIN1* (r_g = 0.56; p-value = 9.80-06) and *APOE* (rg = 0.80; p-value = 1.97e-124) between AD and LBD, and *TMEM175* between PD and LBD (rg = 0.65; p-value = 1.49e-05). Regional genetic correlations between sc-eQTLs and ND highlighted significant correlations (FDR < 0.01) across 33 unique genes, spanning both the innate and adaptive immune systems, including 9 within the Human Leukocyte Antigen (HLA) region correlated with 4 of the tested ND (AD, PD, MS and ALS). Validation of our results with the bulk eQTLGen Consortium dataset (N = 31,684) showed a high correlation between both (cor = 0.73; p-value = 5.68e-10). The outcomes of this study help understand the aetiology of ND, which can be potentially used to repurpose existing immunotherapies used in clinical care for other auto-immune diseases, to slow the progression of ND.

S66. Towards the defeat of neurodegenerative disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 10:30 am - 12:00 noon

ProgNbr 431. Cell-type transcriptome-wide association studies and fine-mapping via deconvolution using single-cell RNA-seq

Authors:

H. Shi[1], M. Zhang[1], K. Siewert[1], S. Gazal[2], A. Pisco[3], A. Nathan[4], S. Raychaudhuri[5], A. Price[1]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Univ. of Southern California, Los Angeles, CA, [3]CZ Biohub, San Francisco, CA, [4]Harvard Univ., Boston, MA, [5]Brigham and Women's Hosp., Boston, MA

Abstract:

Transcriptome-wide association studies (TWAS) integrating gene expression predictions from cis SNPs with GWAS summary statistics have identified thousands of genes associated to disease (Wainberg et al. 2019 *Nat Genet*). However, TWAS generally use gene expression predictions for bulk tissues, and do not leverage fine-grained cell-types.Here, we introduce a powerful approach for performing TWAS and gene-level fine-mapping at cell-type resolution, leveraging large sample sizes for bulk tissues (GTEx Consortium 2020 *Science*) and high-resolution scRNA-seq data (Tabula Sapiens 2022 *Science*; Tabula Muris 2020 *Nature*). We deconvolve the cell-type-specific gene expression of each GTEx sample with respect to Tabula Sapiens and Tabula Muris cell types under an empirical Bayes framework, enabling gene expression prediction and TWAS in each cell type. We also perform gene-level TWAS fine-mapping using a sum of single effects model (analogous to SuSiE; Wang et al. 2020 *JRSSB*) to fine-map causal (gene, cell type) pairs.We applied cell-type TWAS to 52 complex diseases/traits (average N=345K), deconvolving 595 Tabula Sapiens/Tabula Muris cell types using 36 GTEx bulk tissues. Cell-type TWAS identified a median of 42% more independent gene-disease associations per disease than tissue-level TWAS (549 vs. 385), generally in rarer cell types (average cell-type proportion of 8%, vs. 25% for tissue-level TWAS). For example, we identified an association of *CACNA1C* expression in neurons with schizophrenia ($P=7\times10^{-11}$; orders of magnitude stronger than tissue-level TWAS). Although other brain cell types were also significant, TWAS fine-mapping strongly prioritized *CACNA1C* expression in neurons over other cell types (posterior causal probability = 0.97), consistent with known biology. We also identified an association of *APOE* expression in memory B cells with Alzheimer's disease (AD) ($P=1.6\times10^{-6}$; non-significant for tissue-level TWAS). TWAS fine-mapping strongly prioritized *APOE* expression in memory B cells over other cell types (posterior causal probability = 0.99), consistent with the immune basis of AD.We also leveraged cell-type TWAS fine-mapping results to identify disease-associated gene pathways. Cell-type TWAS fine-mapping yielded a median of 62% more independent pathway-disease associations per disease than tissue-level TWAS fine-mapping (44 vs. 26). For example, we identified an association of the pyrimidine metabolism pathway in T helper cells with anorexia ($P=6\times10^{-7}$; non-significant for tissue-level TWAS fine-mapping), a novel finding consistent with the fact that patients with anorexia often have disrupted pyrimidine metabolism.

S67. Context matters! Tissue, cell type, and condition-specificity of epigenetics and expression

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 434. Application of RNA sequencing on transdifferentiated patient fibroblasts for genetic diagnosis of neurological disorders

Authors:

S. Li[1], J. C. Sinson[1], C. Yeh[1], A. Bajic[2], J. A. Rosenfeld[1], J. Sheppard[2], M. Nguyen[2], A. Khoramnia[1], P. P. Hernandez[1], S. C. S. Nagamani[1], Z. liu[2,3], B. Lee[1,4], Undiagnosed Diseases Network, C. Eng[1,5], P. Liu[1,5]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst. at Texas Children's Hosp., Houston, TX, [3]Dept. of Pediatrics, Section of Neurology, Baylor Coll. of Med., Houston, TX, [4]Texas Children's Hosp., Houston, TX, [5]Baylor Genetics, Houston, TX

Abstract:

A prompt and accurate molecular diagnosis is critical for the management of patients with Mendelian disorders. Despite the increasing rates of molecular diagnosis with clinical exome and whole genome sequencing, about half of the patients remain undiagnosed. RNA sequencing (RNA-Seq) has recently been successfully used to detect aberrant transcript events and understand the role of non-coding variants of unknown significance. A major challenge for diagnostic RNA-Seq is that not all genes are well expressed in clinically accessible samples. In a group of 2681 OMIM genes associated with neurological phenotypes, only 63% or 28% are well-expressed [Transcripts Per Million > 10] in fibroblasts or blood, respectively. We assembled a panel of 166 known Mendelian disease genes responsible for brain malformation; 66% were well expressed in fibroblasts and only 26% in blood. To overcome the limitation of tissue-limited expression for diagnostic RNA-Seq, we implemented a protocol optimized for clinical diagnostics for rapid induction of neurons from human fibroblasts. We tested four neuronal induction protocols and obtained similar conversion rates and optimal neuron morphology, with robust expressions of neuron-specific genes. Of genes not well expressed in the brain malformation panel, 34-48% were successfully activated after neuron induction, while 18-25% were boosted for the neurological OMIM group. We are processing more than 60 fibroblasts from patients enrolled in the Baylor College of Medicine Undiagnosed Diseases Network clinical site for neuron induction and RNA-Seq. The selected patients have clinical presentations of brain malformation, intellectual disability, and global developmental delay. We identified diagnostic aberrant splicing events specific to induced neurons that would otherwise have not been detected in fibroblasts (E.g., c.5935-17G>A intronic variant causing a 15bp intronic retention and whole intron 45 retention in *ITPR1*; c.1174+4603G>T deep intronic variant leading to large intronic retention in *DCX*). In addition, we detected aberrant expression in a neuron-specific transcript resulting from a 5' UTR deletion in *MBD5*. As part of the ongoing efforts, we are optimizing the induced neuron RNA-Seq analysis workflow for clinical identification of a broader range of diagnostic variants. In conclusion, we have developed an RNA-Seq analysis workflow based on induced neurons to uncover molecular diagnoses for neurological disorders. The protocol is low cost and has an acceptable turnaround time. Implementation of this protocol improves variant interpretation and molecular diagnosis of undiagnosed Mendelian disorders.

S67. Context matters! Tissue, cell type, and condition-specificity of epigenetics and expression

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 435. Cell-type specific DNA methylome signatures reveal epigenetic mechanisms for neuronal diversity and neurodevelopment disorder

Authors:

Y. Jin, K. Su, H. Kong, W. Ma, Z. Wang, Y. Li, R. Li, E. Allen, H. Wu, P. Jin; Emory Univ., Atlanta, GA

Abstract:

DNA methylation plays critical function in establishing and maintaining cell identity in brain. Disruption of DNA methylation-related processes lead to diverse neurological disorders. However, the role of DNA methylation characteristics in neuronal diversity remains underexplored. Here, we report detailed context-specific DNA methylation maps for GABAergic (GABA), Glutamatergic (Glu) and Purkinje neurons, together with matched transcriptome profiles. Genome-wide mCH levels are distinguishable, while the mCG levels are similar among the three cell types. Substantial CG-DMRs are also seen, with Glu neurons experiencing substantial hypomethylation events. The relationship between mCG levels and gene expression display cell-type specific patterns, while genic CH methylation exhibits a negative effect on transcriptional abundance. We found that cell type-specific CG-DMRs are informative in terms of represented neuronal function. Furthermore, we observed that the identified Glu-specific hypo-DMRs have a high level of consistency with the chromatin accessibility of excitatory neurons and the regions enriched for histone modifications (H3K27ac and H3K4me1) of active enhancers, suggesting their regulatory potential. Hypomethylation regions specific to each cell type are predicted to bind neuron type-specific transcription factors. Finally, we show that the DNA methylation changes in a mouse model of Rett syndrome, a neurodevelopmental disorder caused by the de novo mutations in MECP2, are cell type- and brain region-specific. Our results suggest that cell-type specific DNA methylation signatures are associated with the functional characteristics of the neuronal subtypes. The presented results emphasize the importance of DNA methylation-mediated epigenetic regulation in neuronal diversity and disease.

S67. Context matters! Tissue, cell type, and condition-specificity of epigenetics and expression

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 436. Single-nucleus multiomics reveal enrichment of disease heritability in cell-state-dependent regulatory elements

Authors:

A. Gupta[1,2], T. Amariuta[3], K. Weinand[4], A. Nathan[1], S. Sakaue[4], A. Price[3], L. Donlin[5], K. Wei[6], S. Raychaudhuri[6,2]; [1]Harvard Univ., Cambridge, MA, [2]Broad Inst., Cambridge, MA, [3]Harvard Sch. of Publ. Hlth., Boston, MA, [4]Harvard Med. Sch., Boston, MA, [5]Hosp. for Special Surgery, New York, NY, [6]Brigham and Women's Hosp., Boston, MA

Abstract:

Approximately 90% of disease-associated variants discovered through genome-wide association studies (GWAS) reside in non-coding regions of the genome. For most of these variants, we have yet to define the mechanism by which they influence disease risk. Here, we investigate the extent to which open chromatin regions associated with changes in cellular states account for the heritability of disease risk. We introduce a paradigm that leverages shared, continuous transcriptomic and epigenomic structure at the single-cell level to partition disease heritability. Towards this aim, we integrate multi-modal single-nucleus (sn) snRNA- and snATAC-seq data-28,698 cells from 12 donors across five cell types-with GWAS summary statistics for 28 polygenic diseases. We first defined principal components of gene expression profiles to capture cell-state changes within a cell-type. Using a multivariate Poisson model with the top five PCs, we then identified ATAC peaks that are cell-state dependent or “dynamic” (peaks whose accessibility changes were linearly associated with these expression PCs at likelihood ratio test FDR < 0.05). We find that in T cells, these dynamic peaks are 2.7-fold enriched in promoter regions of T cell-specific genes relative to broadly expressed genes ($p<1e-14$), suggesting a cell-type-specific active transcriptional regulatory role. We recapitulate heritability enrichments in known cell-types; for example, >55% of Rheumatoid Arthritis (RA) and >93% of Celiac’s disease heritability, respectively, is captured by ATAC peaks open in T cells. We further find a >2.3-fold increase in RA heritability enrichment of peaks associated with dynamic T cell-state changes, compared to invariant peaks (29 vs 12.4). We defined a continuous, heritability-based score per cell with which we identify genes, regulatory elements, and cell subtypes associated with genetic risk, suggesting potential therapeutic targets. For example, we find that T follicular helper cells are expanded in autoimmune and inflammatory diseases. Finally, we find that the 14 immune-mediated diseases largely overlap in their heritability-enriched dynamic T cell-states (pairwise Pearson $R \geq 0.8$), highlighting similar disease mechanisms. Overall, this work demonstrates the utility of identifying cell-state-dependent regulatory processes involved in the pathology of polygenic disease.

S67. Context matters! Tissue, cell type, and condition-specificity of epigenetics and expression

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 437. Leveraging human genetic variation and single cell RNA-seq to reveal the catalog of influenza-induced splicing QTLs

Authors:

L. Wang[1], B. Schott[2], D. Ko[2]; [1]Duke Univ, Durham, NC, [2]Duke Univ., Durham, NC

Abstract:

Influenza A virus (IAV) is a substantial burden to global public heath, infecting approximately 30 million people every year and resulting in 300,000 - 650,000 deaths. Of those infected people, symptoms vary from asymptomatic to death. Such variable severity across individuals arises from interaction of genetics and environment. Several recent studies identified a number of single nucleotide polymorphism (SNPs) which are correlated with host gene expression after IAV infection, including our recent work introducing scHi-HOST (single cell High-throughput Human in vitrO Susceptibility Testing) (Schott et al. 2022). While many SNPs are associated with overall gene expression or composite abundance of all transcript isoforms per gene, the extent and molecular mechanisms of how SNPs impact individual transcript isoform usage during IAV infection remain elusive. Here, we investigate how human genetic variants impact alternative isoforms under influenza infection by using single cell RNA-seq (scRNA-seq) and human nature variation. Two independent scRNA-seqs have been conducted from baseline and influenza-infected immortalized B lymphocytes (LCLs), each from a pooling of 48 LCLs representing diverse genetic ancestry from European, Asian and African. By using a customized integrative computational pipeline including read mapping, sample de-convolution, GWAS and splice QTLs (sQTLs) analysis, we uncovered more than 10% of the transcriptome exhibited differential splicing and identified more than 15000 influenza-induced sQTLs gene pairs. Of those sQTLs, more than 40% replicated in GTEx v8. Functional enrichment analysis of the genes harbouring the sQTLs highlighted the role of interferon related pathways. We further performed whole genome sequencing and time-course bulk RNA-seq in a human flu challenge cohort, and integrated sQTLs and splicing genes identified from scRNA-seq in cells with those identified in infected human volunteers. Collectively, we identified and validated a catalog of influenza-induced sQTLs and splicing genes that highlights the role of host transcript usage driven by the interaction of host genetics and IAV infection. This study expands our work of scHi-HOST and suggests alternative splicing plays a critical role in influenza infection, advancing our understanding of the genetic architecture of flu susceptibility and facilitating future diagnostics and treatments.

S67. Context matters! Tissue, cell type, and condition-specificity of epigenetics and expression

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 438. Cell type specific and disease associated eQTLs in the human lung

Authors:

N. Banovich[1], **H. Natri**[2], C. Azodi[3], L. Peter[4], C. Taylor[5], R. Kendle[4], J. Kropski[5], D. McCarthy[3]; [1]TGen, Phoenix, AZ, [2]Translational Genomics Res. Inst. (TGen), Phoenix, AZ, [3]St. Vincent's Inst. of Med. Res., Melbourne, Australia, [4]Translational Genomics Res. Inst., Phoenix, AZ, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN

Abstract:

Understanding the genetic architecture of gene regulation and disease is the main goal of functional genomics, however, it has become increasingly clear that to fully realize the promise of these approaches it is necessary to assay the genetic control of gene regulation within the proper context - both the right cell-type and under the proper conditions. Here we apply these principles to map eQTLs in the human lung across health and disease. Specifically, our efforts are focused on understanding the genetic architecture of interstitial lung disease (ILD). ILD is a chronic, progressive lung disease characterized by the scarring of lung tissue through epithelial remodeling and accumulation of extracellular matrix (ECM). Pulmonary Fibrosis (PF) is a clinical phenotype that exhibits the end stage of ILD. To enable cell-type specific eQTL mapping from primary human lung tissue we have employed single cell RNA-seq (scRNA-seq) to generate expression profiles from over 400,000 cells of 116 individuals (49 healthy and 67 ILD). In accordance with previous work from our group defining best practices for single cell eQTL mapping (sc-eQTL) we are using pseudo-bulk expression levels, aggregating reads across all cells of the same cell-type from an individual. With this approach, we are able to identify a number of sc-eQTLs across cell-types. As expected, we find our power to detect sc-eQTLs is driven in large part by overall cell-type abundance, with cell-types present at lower frequencies giving rise to fewer sc-eQTLs. Given the differences in power between cell-types, we used the multivariate adaptive shrinkage (mashr) approach to identify shared and cell-type specific eQTLs. Using this approach, we detect clear cell population, lineage, and cell-type-specific signals. To identify disease specific sc-eQTLs we mapped eQTLs separately in PF case and healthy control individuals and applied mashr to estimate shared and disease specific effects. Finally, we have intersected our results with recent genome-wide association studies (GWAS) of ILD. We identify cell-type specific eQTLs that colocalize with IPF GWAS risk loci. These include a well-characterized risk locus near the DSP gene that confers an IPF risk odds ratio of 1.44 which we found to be an eQTL in alveolar cells. This work aids in determining the cell types and states in which PF-associated genetic variants function, and highlights the importance of cellular context in gene regulation in health and disease.

S67. Context matters! Tissue, cell type, and condition-specificity of epigenetics and expression

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 439. Brain cell-type-specific protein interactomes of rare variant schizophrenia risk genes highlight shared biology with neurodevelopmental disorders

Authors:

Y-H. Hsu[1], G. Pintacuda[1], J. Biagini[1], J. Ching[1], N. Fornelos[1,2], K. Lage[1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA

Abstract:

Large-scale genetic studies such as the schizophrenia exome meta-analysis (SCHEMA) have nominated many schizophrenia risk genes and implicated the importance of specific brain tissues and cell types in the disease. However, we often lack functional interpretation for the identified risk genes, especially in the context of the disease-relevant cell types. To empower such interpretation, we performed interaction proteomics for 10 schizophrenia risk genes prioritized by rare variant associations in SCHEMA in human iPSC-derived neural progenitor cells (NPCs) and induced excitatory neurons (iNs). The resulting protein-protein interaction (PPI) networks contain > 86% newly reported PPIs, capture cell-type-specific biology in NPCs vs. iNs, and reflect gene relationships in complex brain tissues as measured by multiple independent transcriptomic datasets. Interestingly, while both the NPC-specific and iN-specific networks are enriched for genetic risks of severe developmental disorders, only the iN-specific network is enriched for genetic risks of schizophrenia and autism spectrum disorders. These results not only agree with genetic studies that observed convergent biology between schizophrenia and neurodevelopmental disorders, but also provide clues on how schizophrenia risk genes may contribute to these phenotypes through interacting with different proteins at distinct developmental stages. Finally, we intersected the networks with exome sequencing data to prioritize additional sub-threshold schizophrenia risk genes, and with GWAS data to complement conventional approaches such as statistical fine-mapping and eQTL co-localization to nominate causal genes within GWAS loci. Overall, our findings support the known genetic correlations between schizophrenia and neurodevelopmental disorders, provide insights into the underlying molecular networks at different stages of neuronal maturation, and showcase brain cell-type-specific PPI data as a systematic framework for translating genetic data into actionable biological hypotheses for studying these disorders. In the future, this framework can be applied to study a wider range of risk genes and cell types as larger genetic datasets and optimized cell modeling protocols become increasingly available.

S68. Leveraging population genetics to inform diverse cohorts and biobanks

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 442. Advancing fine-scale population health monitoring systems in a Los Angeles biobank

Authors:

C. Caggiano[1], A. Boudaie[2], R. Shemirani[3], E. Petter[1], A. Chiu[4], R. Johnson[1], D. Ercelen[1], B. Pasaniuc[5], E. Kenny[3], J. Shortt[6], C. Gignoux[6], B. Balliu[1], V. Arboleda[1], G. Belbin[7], N. Zaitlen[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Oscar Hlth.Inc, New York, NY, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Univ. of California Los Angeles, Los Angeles, CA, [5]Geffen Sch. of Med. at UCLA, Los Angeles, CA, [6]Univ. of Colorado Denver Anschutz Med. Campus, Aurora, CO, [7]Gencove Inc., New York, NY

Abstract:

Background: The population that an individual belongs to influences their risk for disease, whether through a shared environment or through genetics. The study of fine-scale populations in clinical care is important for reducing health disparities and for developing personalized treatments. Los Angeles is an ideal backdrop for studying population health, as it is characterized by population diversity stemming from both recent and historical immigration. **Methods:** In this work, we developed a health monitoring system, which leverages biobank data and electronic medical records from 40,000 diverse UCLA patients. We identified 376 LA communities using identical by descent (IBD) segments and examined their interactions with the UCLA Health system. These included many understudied populations, like communities of Lebanese Christians, Iranian Jews, Armenians, and Gujaratis. **Results:** Our analyses identified thousands of novel associations between IBD communities and clinical diagnoses, including higher odds of having a heart transplant in the Armenian IBD community (O.R.: 4.01±1.65, p=4.72e-08) and higher odds of adjustment disorder in the Iranian Jewish IBD community (O.R: 2.89±1.42, p-value: 2.04e-09). Furthermore, we observed how the rates of diagnosis of disease codes can change with time. For example, rates of kidney transplants significantly increased between 2016 and 2020 in the Central American IBD community (p= 3.79e-09). To understand how communities access care in LA on a larger scale, we also examined what offices and specialties IBD communities visited. The European Non-Jewish community was significantly associated with visiting a primary care office (O.R: 1.32±1.05, p=6.34e-30), while many non-European IBD communities were less likely to visit specialties associated with preventative care. Lastly, we sought to understand population genetic risk. We found exceptionally high IBD sharing within the Iranian Jewish (n=264) community (mean=57.43 cM), suggesting a potential founder effect that could have implications for inherited disease testing. We also identified elevated rates of pathogenic alleles in IBD communities, such as elevated MAF of an allele associated with alpha-thalassemia in the Iranian Non-Jewish community (p=5.02e-05). **Conclusions:** This work represents an advance toward equitable health research and can empower future studies on population health in Los Angeles. To enhance the impact of the research we provide a web portal of our results, including disease associations, genetic summary statistics, and an interactive map of where communities are most likely to access care https://christacaggiano.github.io/ibd

S68. Leveraging population genetics to inform diverse cohorts and biobanks

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 443. A comprehensive genetic profile of 140,000 adults from the Mexico City Prospective Study

Authors:

J. Torres[1,2], A. Ziyatdinov[3], J. Alegre-Díaz[4], J. Backman[3], J. Mbatchou[3], M. Turner[1,5], S. Gaynor[3], T. Joseph[3], Y. Zou[3], D. Liu[3], R. Wade[1,2], J. Staples[3], A. Popov[3], X. Bai[3], S. Balasubramanian[3], L. Habegger[3], E. Maxwell[3], M. Jones[3], H. García-Ortiz[6], R. Ramirez-Reyes[4], R. Santacruz-Benítez[4], A. Nag[7], K. Smith[7], Regeneron Genetics Center, M. Reppell[8], S. Zöllner[9], E. Jorgenson[3], W. Salerno[3], S. Petrovski[7], J. Overton[3], J. Reid[3], T. Thornton[3], G. Abecasis[3], L. Orozco[6], R. Collins[1], A. Baras[3], M. R. Hill[1], J. Berumen[4], J. Emberson[1], J. Marchini[3], P. Kuri-Morales[4], R. Tapia-Conyer[4]; [1]Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [2]MRC Population Hlth.Res. Unit, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [3]Regeneron Genetics Ctr., Tarrytown, NY, [4]Experimental Res. Unit from the Faculty of Med. (UIME), Natl. Autonomous Univ. of Mexico (UNAM), Mexico City, Mexico, [5]Oxford Kidney Unit, Churchill Hosp., Oxford, United Kingdom, [6]Natl. Inst. of Genomic Med. (INMEGEN), Mexico City, Mexico, [7]Ctr. for Genomics Res., Discovery Sci., Res. and Dev. Biopharmaceuticals, AstraZeneca, Cambridge, United Kingdom, [8]AbbVie Inc, North Chicago, IL, [9]Dept. of Biostatistics, Univ. of Michigan, Ann Arbor, MI

Abstract:

Latin American populations remain largely underrepresented in human genetics research. To address this disparity, we performed a comprehensive genetic profile of 140,000 adults from the Mexico City Prospective Study (MCPS), the largest blood-based study in a Latin American population. Between 1998 and 2004, participants were recruited from two urban districts in Mexico City (Coyoacán and Iztapalapa), each containing a diverse mix of long-term residents and recent migrants. Genotyping and whole exome sequencing (WES) of the cohort, and whole genome sequencing (WGS) of 10K individuals, was performed by the Regeneron Genetics Center. We estimated genetic relatedness from shared identity-by-descent (IBD) segments and found that 71% of participants had at least one relative in the study that was third-degree or closer, with many participants having multiple close relatives. Strikingly, the largest connected component of a graph of third-degree relationships included 30,682 individuals (22% of the cohort). Moreover, we used PRIMUS to reconstruct 22,766 first-degree family networks that involved 3,595 nuclear families. By incorporating external reference samples, including 716 indigenous samples from the Metabolic Analysis of an Indigenous Sample (MAIS) study representing 60 of the 68 recognized ethnic groups in Mexico, we characterized population structure and ancestry components within MCPS. Through a series of principal components analyses, parametric admixture estimation, and fine-scale haplotype-based methods, we elucidated complex patterns of Native American, European, and African ancestry and resolved extensive Mesoamerican admixture from central, southern, and southeastern Mexico. Local ancestry inference further revealed an excess of African ancestry at the MHC locus on chromosome 6 and elevated Native Mexican ancestry on chromosome X with estimated female contributions of 71.3% and 7.5% for Native Mexican and European ancestries, respectively. Lastly, an WES variant survey identified an excess of homozygous predicted loss-of-function variants in MCPS compared to the UK Biobank WES dataset. We delineated runs of homozygosity (ROH) from phased genotypes and determined that 60,722 participants (44% of the cohort) had at least one ROH segment greater than or equal to 4 cM. Incorporating local ancestry showed that 79% of ROH segments were assigned to Native Mexican ancestry, which collectively harbored 64% of all rare homozygous pLOF variants. Overall, the complex patterns of relatedness, admixture, and variation in the MCPS cohort enhances the genetic diversity uncovered from Latin American genomes.

S68. Leveraging population genetics to inform diverse cohorts and biobanks

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 444. Analysis of a large cohort of admixed Greenlandic siblings shows that genetic load of metabolic phenotypes differs between Inuit and Europeans

Authors:

L. Lin[1], F. Stæger[1], K. Hanghøj[1], Z. Li[1], C. Larsen[2], M. Jørgensen[3,4], N. Grarup[5], I. Moltke[1], T. Hansen[5], A. Albrechtsen[1]; [1]Section for Computational and RNA Biology, Dept. of Biology, Univ. of Copenhagen, Copenhagen, Denmark, [2]Ctr. for Publ. Hlth.Res. in Greenland, Natl. Inst. of Publ. Hlth., Univ. of Southern Denmark, Copenhagen, Denmark, [3]Clinical Res., Copenhagen Univ. Hosp. – Steno Diabetes Ctr. Copenhagen, Herlev, Denmark, [4]Ctr. for Publ. Hlth.Res. in Greenland, Natl. Inst. of Publ. Hlth., Univ. of Southern Denmark. Steno Diabetes Ctr. Greenland, Nuuk, Greenland, [5]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark

Abstract:

The genetic load of a phenotype can differ between populations depending on their genetic background, meaning that they might have different average trait values solely due to genetics. Genetic load is important for understanding disease prevalence, relative heritability and fitness in evolution among different populations. However, genetic load cannot be estimated directly because we do not know all causal variants and their effect size, and differences in genetic load between human populations are challenging to quantify because environmental factors also vary between populations.Here we use a unique study design to estimate genetic load differences between Inuit and Europeans by using genetically admixed full sibling pairs. There are small differences in ancestry between admixed siblings due to random recombination. We use linear regression to correlate the sib pairs difference in ancestry proportions with their difference in phenotypes. The regression slope is an estimator of the difference in genetic load. This design has the strength that it is robust to differences in environment between families of different genetic ancestry. Using dense genotype data of 4,607 Greenlandic individuals, we inferred 1,339 full sibling pairs, where both siblings are admixed with Inuit and European ancestry. We analysed 31 metabolic traits and found 9 of them having significant difference in genetic load between European and Inuit genetic ancestry. Most of the differences exist in body composition, where Inuit have a lower genetic load, including lower weight, waist as well as hip and lean mass. This reveals that the observed difference in body composition between Greenlanders and Europeans is partly explained by their genetics. This is the first time genetic load differences have been shown to exist between human populations using a method which is not confounded by possible differences in environmental factors.

S68. Leveraging population genetics to inform diverse cohorts and biobanks

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 445. Genomic analysis of 15,154 individuals from India: A largest study to-date to explore understudied Indian sub-populations

Authors:

M. Kapoor[1], B. Ye[1], J. Backman[1], S. Tuna[2], N. England[2], O. Babu[2], J. Rodriguez-Flores[1], T. Joseph[1], A. Ziyatdinov[1], V. Rajagopal[3], S. Jamuar[4], D. Liu[1], W. McLaren[2], N. Sarkar[2], A. Shah[2], K. Sharma[2], G. Tzoneva[1], J. Marchini[1], Global Gene Corporation, Regeneron Genetics Center, J. Picker[2], A. Shuldiner[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Global Gene Corp. PTE., Ltd., Singapore, Singapore, [3]Regeneron Genetics Ctr., White Plains, NY, [4]KK Women's and Children s Hosp., Boston, MA

Abstract:

Introduction: Imbalance in availability of large genomic data in diverse global populations limits our understanding of complex genetic architecture of worldwide populations. Here, we report results from the largest study of coding variants obtained from whole exome sequencing of more than 15K individuals from five large states (Rajasthan, Maharashtra, Uttar Pradesh, Gujarat and Karnataka) of India. Past, much smaller genetic studies on South Asian Indian populations focused on a few selected subgroups of ethnically diverse populations to explore genetic admixture and migration patterns within geographical regions. In contrast, our analysis is an effort to tease apart the genetic differences in linguistic and socio-cultural diversity of all major subgroups residing within different geographical regions of India. **Methodology:** Unrelated subjects (N = 15,154) were recruited as part of an ongoing collaboration between the Regeneron Genetics Center and Global Gene Corporation to study genetic diversity across India. Genotyping was performed using Illumina GSA-24v2-0_A2 arrays and whole exome sequencing. **Results:** Using principal components and Uniform Manifold Approximation and Projection (UMAP) based analyses we mapped each Indian sub-population to distinct groups or clusters which were highly correlated with self-reported ethnicities. Unsupervised admixture analysis further demonstrated the increased genetic diversity and differentiation in sub-populations from Northern to Southern parts of India. We also conducted a genome-wide scan of selection signatures by estimating locus-specific time to the most recent common ancestor using ascertained sequentially Markovian coalescence (ASMC). The analysis identified 24 (14 known: 10 novel) genetic loci with selection signatures satisfying genome-wide significance. Phenome-wide association analysis of variants and genes under selected loci (in UKB-all, UKB-South Asians, current dataset) indicated associations with a variety of complex traits involved in adaptation of South Asian populations. **Conclusion:** We conducted the largest study to date to tease apart the linguistic, cultural and geographical differences that shape genetic diversity in India. Our study identified unique population-specific features and novel selection signatures in Indian sub-populations, highlighting the value of diverse data in genomic studies.

S68. Leveraging population genetics to inform diverse cohorts and biobanks

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 446. Integrative functional genomic analyses identify novel genetic variants influencing skin pigmentation in Africans

Authors:

Y. Feng[1], F. Inoue[2,3], S. Fan[1,4], N. Xie[1], C. Zhang[1], F. Zhang[1], D. Kelly[1], M. Hansen[1], E. Oancea[5], M. S. Marks[6], T. Nyambo[7], S. Mpoloka[8], G. G. Mokone[9], A. Njamnshi[10], C. Folkunang[11], G. Belay[12], N. Ahituv[3,13], S. Tishkoff[1]; [1]Dept. of Genetics and Biology, Univ. of Pennsylvania, Philadelphia, PA, [2]Inst. for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto Univ., Kyoto, Japan, [3]Dept. of Bioengineering and Therapeutic Sci., Univ. of California San Francisco, San Francisco, CA, [4]Human Phenome Inst., Sch. of Life Sci., Fudan Univ., Shanghai, China, [5]Dept. of Molecular Pharmacology, Physiology and Biotechnology, Brown Univ., Providence, RI, [6]Dept. of Pathology and Lab. Med., Children's Hosp. of Philadelphia Res. Inst., Philadelphia, PA, [7]Dept. of Biochemistry, Kampala Intl. Univ., Dares Salaam, Tanzania, United Republic of, [8]Dept. of Biological Sci., Faculty of Sci., Univ. of Botswana, Gaborone, Botswana, [9]Dept. of BioMed. Sci., Univ. of Botswana, Gaborone, Botswana, [10]Dept. of Neurology, Central Hosp. Yaoundé; Brain Res. Africa Initiative (BRAIN), NeuroSci. Lab, Faculty of Med. and BioMed. Sci., The Univ. of Yaoundé I, Yaoundé, Cameroon, [11]Dept. of Pharmacotoxicology and Pharmacokinetics, Faculty of Med. and BioMed. Sci., The Univ. of Yaoundé I, Yaoundé, Cameroon, [12]Dept. of Biology, Addis Ababa Univ., Addis Ababa, Ethiopia, [13]Inst. for Human Genetics, Univ. of California San Francisco, San Francisco, CA

Abstract:

Variation in human skin color is hypothesized to reflect adaptation to variable solar ultraviolet radiation during modern global human migration from an African origin across the globe. Genetic variants associated with human skin pigmentation have been identified using genome-wide association studies (GWAS), but most of these studies focused on lightly pigmented European populations and little is known about the molecular mechanism underlying skin color diversity in Africans. Using a novel high-coverage whole genome sequencing dataset from 180 individuals originating from twelve ethnically and geographically diverse African populations, we performed F_{ST}-based analysis to detect loci that are highly differentiated between African populations with light (the San population who speak Khoesan languages from Southern Africa) versus dark skin color (other Africans). Additionally, we identified candidate genetic variants associated with skin pigmentation in 1544 Africans by GWAS using 32,574,188 single nucleotide polymorphisms (SNPs) imputed from Illumina Omni 5M array data. We applied a massively parallel reporter assay (MPRA) to screen 1,157 variants identified from these analyses that are located in open chromatin regions in melanocytes and identified 165 SNPs showing significant differential regulatory activities between alleles. We constructed a high-resolution chromatin interaction map of melanocyte-derived cells using Hi-C and H3K27ac HiChIP to identify target genes interacting with candidate functional regulatory SNPs. We validated candidate enhancers using luciferase reporter assays, validated their interaction with target genes (near *MFSD12*, *OCA2, DDB1* and *MITF*) using CRISPR/Cas9-based genome editing, examined their impact on gene expression using RNA-seq, and tested their effects on pigmentation levels in melanoma cells. We detected signals of local adaptation at the *MITF* locus in the San and identified a San-specific variant affecting the enhancer activity of *MITF,* which may contribute to their relatively light skin color. We identified *CYB561A3* as a novel gene impacting pigmentation *in vitro* by impacting genes involved in oxidative phosphorylation and melanogenesis. Together, our study identified novel SNPs, enhancers, and genes involved in skin color variations in African populations and sheds light on the complex genetic mechanisms underlying human skin color evolution.

S68. Leveraging population genetics to inform diverse cohorts and biobanks

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 447. Multiadaptive shrinkage improves cross-population transcriptome prediction for transcriptome-wide association studies in underrepresented populations

Authors:

D. Araújo[1], C. Nguyen[1], X. Hu[2], S. S. Rich[2], J. I. Rotter[3], H. Im[4], A. W. Manichaikul[2], H. E. Wheeler[1], NHLBI TOPMed Consortium; [1]Loyola Univ. Chicago, Chicago, IL, [2]Univ. of Virginia, Charlottesville, VA, [3]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [4]Univ. of Chicago, Chicago, IL

Abstract:

The majority of genome- and transcriptome-wide association studies (GWAS/TWAS) are conducted in European ancestry populations, and consequently may not reflect genetic variants and linkage disequilibrium patterns found within non-European populations. Population-matched TWAS have been shown to increase the number of true discoveries; however, building accurate transcriptome prediction models for underrepresented populations is not always possible due to smaller samples sizes or data availability. Thus, we sought to build new transcriptome prediction models with better cross-population performance.
We used RNA-Seq data from the Trans-omics for Precision Medicine (TOPMed) Multi-Ethnic Study of Atherosclerosis (MESA) from up to 1287 African American (AFA), Chinese (CHN), European (EUR) and Hispanic/Latino (HIS) individuals of 3 blood cell types: peripheral blood mononuclear cells, CD16+ monocytes, and CD4+ T cells. Multivariate adaptive shrinkage in R (MASHR) was developed to estimate effects in genomic studies with multiple conditions and adapts to patterns present in the data, allowing for both shared and condition-specific effects. Here, we first used Matrix eQTL to perform cis-eQTL mapping in each population-tissue pair to generate single population effect sizes and standard errors, which were used as the input for MASHR. We then carried the most significant SNP in each population forward to prediction models, using the MASHR-generated effect sizes. We compared prediction performance in Geuvadis of MASHR models to unadjusted top SNP (Matrix eQTL) and elastic net (EN) approaches. MASHR models performed better than Matrix eQTL in all cases, whereas in comparison to EN, MASHR either performed the same or better. For example, AFA MASHR and EN models performed the same in Geuvadis Yoruba (YRI, p=0.21), but AFA MASHR models performed better than AFA EN in Geuvadis British (GBR, p=2.0x10-24). With the EUR models, MASHR outperformed EN in YRI (p=6.0x10-34) and in GBR (p=0.00052).
Next, we performed TWAS using GWAS summary statistics for 28 complex traits from the Population Architecture using Genomics and Epidemiology study. Across all population models and phenotypes tested, MASHR identified the most significant gene-trait pairs (MASHR=310, Matrix eQTL=295, EN=200), but the least unique significant pairs (MASHR=108, Matrix eQTL=136, EN=132), which could indicate that MASHR models give more consistent results across populations. We expect that by improving cross-population transcriptome prediction, we will identify new, more consistent gene-trait associations to better understand the underlying mechanisms of complex traits.

S69. Methods and databases: Open, benchmarked and FAIR

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 450. Comparing open-source tools to detect alternative splicing and gene expression outliers in RNA-seq to improve diagnostic yield

Authors:

C. Jaramillo Oquendo[1], A. G. L. Douglas[1,2], J. Lord[1], H. Wai[1], D. Baralle[1]; [1]Univ. of Southampton, Faculty of Med., Southampton, United Kingdom, [2]Oxford Univeristy Hosp. NHS foundation Trust, Oxford, United Kingdom

Abstract:

Introduction: RNA sequencing (RNA-seq) is now being used as a complementary tool to DNA sequencing in diagnostics of rare disease where DNA has been uninformative. RNA-seq allows us to identify alternative splicing (AS) and aberrant gene expression allowing for improved interpretation of variants of unknown significance (VUSs). Our aim is to improve diagnostic yield of rare disease using blood-based RNA-seq. An initial step is to identify the best methodology to detect aberrant events in RNA-seq particularly when there are no candidate VUSs. Here we compare open-source tools to detect AS and gene expression outliers in RNA-seq and investigate feasibility to generate new diagnostic candidates using these tools.
Methods: RNA was extracted from blood in 76 patients with likely genetic disorder where 40 had a candidate VUS and 9 had previously identified microdeletions. Samples were sequenced in four batches with 150 bp paired end reads per sample. STAR was used to align to reference genome (GRCh38), HTSeq and Salmon were used to quantify gene and transcript counts. rMATS, MAJIQ and OUTRIDER were used to detect AS and gene expression outliers.
Results: Mean number of sequencing reads was 76.6 million (61.3-130.2 million) and on average 80% of reads were uniquely mapping. The splicing tools rMATS and MAJIQ, each identified an average of 2657, and 797 significant AS events respectively. OUTRIDER identified 289 outliers across 60 samples. Visual inspection of the bam files (IGV) allowed the detection of alternative splicing in 43% (n=17) of samples with a candidate VUSs. 15 of the 17 alternative splicing events were identified by rMATS and only 8 were picked up by MAJIQ. For those where alternative splicing was not detected (n=23), 11 variants could not be assessed due to low gene coverage while the rest showed no splicing abnormalities in the candidate gene. OUTRIDER picked up 50% of known microdeletions and identified candidate genes in samples with no VUSs.
Conclusions: Open-source tools such as STAR, rMATS and MAJIQ can pick up large numbers of AS events in RNA-seq but distinguishing significant events from noise is complicated without a candidate VUS. Identification of splicing and expression abnormalities were limited by gene expression in blood, ultimately affecting gene coverage. Regardless, blood-based RNA-seq allowed us to increase the molecular diagnostic yield by validating splicing abnormalities in 17 patients with a candidate VUS and microdeletions in 5 patients. The next steps will be to create filtering thresholds to inspect those that remain with no candidate VUSs and apply to large cohort (6000+) within the 100,000 Genomes Project.

S69. Methods and databases: Open, benchmarked and FAIR

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 451. GestaltMatcher Database: A FAIR database for medical imaging data of rare diseases

Authors:

H. Lesmann[1,2], S. Moosa[3], S. Rosnev[2], J-T. Pantel[2], A. Hustinx[2], B. Javanmardi[2], A. Knaus[2], T. Kamphans[4], W. Meiswinkel[4], J-M. Li[2], M. Caro Martin[5], C. Velmans[6], M. Höller[2], I. Abdelrazek[7], F. Ebstein[8], S. KÜRY[9], E. Abdalla[7], F. Marbach[5], C. Netzer[6], M. Nöthen[1], G. Lyon[10,11], P. M. Krawitz[2], T-C. Hsieh[2]; [1]Inst. of Human Genetics, Univ. Bonn, Bonn, Germany, [2]Inst. for Genomic Statistics and Bioinformatics, Univ. Bonn, Bonn, Germany, [3]Div. of Molecular Biology and Human Genetics, Stellenbosch Univ. and Med. Genetics, Stellenbosch, South Africa, [4]GeneTalk, Bonn, Germany, [5]Inst. for Human Genetics, Univ. Hosp. Heidelberg, Heidelberg, Germany, [6]Inst. of Human Genetics, Cologne, Germany, [7]Dept. of Human Genetics Med. Res. Inst. Alexandria Univ., Alexandria, Egypt, [8]Inst. für Medizinische Biochemie und Molekularbiologie (IMBM), Univ.smedizin Greifswald, Greifswald, Germany, [9]Inst. de Biologie, CHU de Nantes Hôtel-Dieu; l'Inst. du Thorax, INSERM, CNRS, Université de Nantes, NANTES, France, [10]Dept. of Human Genetics and George A. Jervis Clinic, NYS Inst. for Basic Res. in Dev.al Disabilities, Staten Island, NY, NY, [11]Biology PhD Program, The Graduate Ctr., The City Univ. of New York, New York, NY

Abstract:

Introduction: The value of computer-assisted image analysis has been shown in several studies. The performance of tools with artificial intelligence (AI) such as GestaltMatcher (GM) improves with the size and diversity of the training set. Properly labelled training data is currently the biggest bottleneck in developing next-generation phenotyping (NGP) applications. Therefore, we developed together with medRxiv GestaltMatcher Database (GMDB) - a database for machine-readable medical image data that complies with the FAIR-principles and improves the openness and accessibility of scientific findings in Medical Genetics.
Methods: An entry in GMDB consists of a medical image such as a portrait, X-ray or fundoscopy, and machine-readable meta information such as a clinical feature encoded in HPO terminology or a disease-causing mutation reported in HGVS format. In the beginning, data was mainly collected by curators gathering images from the literature. Now, mainly clinicians and individuals, recruited from patient support groups, provide their previously unpublished data. For this patient-centered approach, we developed a digital consent form. GMDB is a modern publication medium for case reports that complements preprints e.g. on medRxiv. To enable inter-cohort comparisons, we implemented a research feature in GMDB that computes the pairwise syndromic similarity between hand-picked cases.
Results: We compiled an image collection of over 6800 cases with more than 680 disorders in GMDB by a community-driven effort. Most of the data was collected from over 2200 publications. In addition, about 200 previously unpublished cases were obtained. The web interface enables gene- and phenotype-centered queries or infinite scroll in the gallery. The digital consent has led to increasing adoption of the approach by patients. The research app within GMDB was used to generate syndromic similarity matrices to characterize two novel phenotypes (*CSNK2B, PSMC3*).
Conclusion: GMDB is the first FAIR database for NGP where data are findable, accessible, interoperable, and reusable. It serves as a repository for medical images that cannot be included in medRxiv. By that means GMDB connects clinicians with a shared interest in particular phenotypes and improves the performance of AI.

S69. Methods and databases: Open, benchmarked and FAIR

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 452. Polygenic risk score prediction accuracy: A retrospective analysis

Authors:

H. Aschard[1], J. Kim[2], H. Julienne[1], C. Albiñana[3], C. Boetto[1], B. Vilhjálmsson[3], F. Privé[3], L. Henches[1]; [1]Inst. Pasteur, Paris, France, [2]Harvard T. H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Aarhus Univ., Aarhus, Denmark

Abstract:

Polygenic risk score (PRS) based on thousands of genetic variants has become a central tool for genetic prediction in multifactorial diseases, and the prospect of using these PRS in clinical care has received increasing attention. The accuracy of PRS will likely continue to increase with increasing sample size of genome-wide association (GWAS) and the development of new powerful methods. However, questions remain on how much prediction can be achieved in the future and how performances depend on the many parameters involved. Here, we used both real data and theoretical models to provide both a retrospective on PRS performances and a perspective on future prediction accuracy.
First, we curated over 30 GWASs from six common diseases published since 2006 and including 1K to 130K cases, and for each of them derived PRS using a harmonized processing pipeline. We applied those PRSs to independent samples and show that while the prediction accuracy, as measured by the AUC, tends to converge toward a maximum as a function of the sample size for some outcomes (type 2 diabetes, breast cancer), the AUC for other outcomes (asthma, coronary artery disease⋯) did not show any clear trend. This heterogeneity appears to be driven by variability in the case definition in the original GWAS (age at onset, self-reported vs doctor diagnosed, etc), suggesting that future progresses will not only rely on sample size, but also on careful phenotyping.
We then investigated the expected maximum prediction accuracy conditional on multiple parameters. We used the UK Biobank imputed data to derived the proportion of heritability that can be captured by both genotyped and imputed variants across various genetic effect distribution based on the so-called alpha-model. We show that in many scenarios, variants genotyped and those with high imputation quality only capture a fraction of the total heritability, limiting the maximum achievable AUC, and suggesting that reaching the maximum accuracy will mostly rely on including rare variants in PRSs.
Finally, we demonstrate that the fit of the alpha model on the six diseases can be greatly improved by applying ad hoc corrections on the effect size distribution. Using these optimized models, we derived updated trends in AUC as a function of sample size for each disease, providing a robust perspective on the convergence toward the maximum.

S69. Methods and databases: Open, benchmarked and FAIR

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 453. The ENCODE 4 long-read RNA-seq resource reveals distinct classes of isoform diversity

Authors:

F. Reese[1], B. A. Williams[2], M. H. Çelik[1], B. Borsari[3], K. Bedi[4], D. Trout[2], E. Rebboah[1], ENCODE RNA Working Group, M. Ljungman[4], R. Guigó[5], B. J. Wold[2], A. Mortazavi[1]; [1]Univ. of California, Irvine, Irvine, CA, [2]California Inst. of Technology, Pasadena, CA, [3]Yale Univ., New Haven, CT, [4]Univ. of Michigan, Ann Arbor, MI, [5]Ctr. for Genomic Regulation, Barcelona, Spain

Abstract:

A significant proportion of mammalian genes encode for multiple transcript isoforms that result from differential promoter usage, changes in internal splicing, and 3' end choice. The comprehensive characterization of transcript diversity across tissues, cell types, and species has been challenging because transcripts are much longer than reads normally used for RNA-seq. Long-read RNA-seq (lrRNA-seq) allows for identification of the complete structure of each transcript. As part of the final phase of the ENCODE Consortium, we sequenced 216 lrRNA-seq libraries totalling 1 billion circular consensus reads (CCS) for 60 unique human and mouse samples. We detected and quantified 94.4% of GENCODE protein coding genes as well as 42.6% of known protein coding transcripts. Overall, we detected over 100,000 full-length transcripts, one third of which are novel. We then define a new reference set of transcription start sites (TSSs), transcription end sites (TESs), and intron chains that are used for each gene across diverse tissues and cell types. Finally, we develop new metrics to characterise the transcriptional diversity of each gene in terms of alternative TSS choice, TES choice, and internal splicing; and demonstrate that this diversity varies on a per-gene basis across tissues, cell lines, and species. Our results represent the first comprehensive survey of human and mouse transcriptomes using full-length long reads and will serve as a foundation for further transcript-centric analyses.

S69. Methods and databases: Open, benchmarked and FAIR

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 454. Evaluating gene prioritization methods and their ability to identify successful drug targets

Authors:

M. Sadler[1], P. Deelen[2], Z. Kutalik[3]; [1]Univ. Ctr. for Primary Care and Publ. Hlth., Lausanne, Switzerland, [2]Univ. of Groningen, Univ. Med. Ctr. Groningen, Dept. of Genetics, Groningen, Netherlands, [3]Univ. of Lausanne, Lausanne, Switzerland

Abstract:

Drugs whose targets have genetic support were found to be more likely to succeed in trials. Gene-based genome-wide association studies (GWAS), rare variant burden tests in whole exome sequencing (WES) studies or integration of GWAS with quantitative trait loci (QTL) data have been proposed to identify likely causal genes for complex diseases. Here, we compared several such gene-prioritization approaches on 20 common clinical traits and benchmarked their ability in recovering genes targeted by drugs prescribed to treat these medical conditions (according to the DGIdb and STITCH databases). We found that gene scoring methods that map GWAS signals to genes based on proximity and local linkage disequilibrium structure (Pascal) achieved the highest performance (mean OR=2.05 for top 1% prioritized genes across traits). Genes prioritized by burden tests that aggregate rare variants (MAF $\leq$ 1%) from WES data of the UK Biobank exhibited overall inferior performance (OR=1.51, P_{diff}=5.8e-3) compared to Pascal, still predicting well targets of cholesterol-lowering medications (OR=7.3 (P=1.7e-9)). Although mechanistically most insightful, genes ranked top in Mendelian randomization (MR) analyses combining expression QTLs (GTEx project; eQTLGen consortium) and GWAS data, showed limited enrichment with drug target genes (overall OR=1.45 with highest enrichment (OR=4.2, P=0.02) for chronic kidney disease). The poorer performance may be the consequence of drug target transcript/protein levels being more polygenic and less heritable (based on eQTL (eQTLGen, N=31,684) and pQTL (deCODE genetics, N=35,559) data), suggesting that they are under tighter regulatory constraints and less amenable to MR-based approaches. Furthermore, our analysis showed that diffusing the GWAS-identified genes (by Pascal) on a protein-protein interaction network (STRING) massively boosts overall enrichment to OR=6.63. However, this improvement may be due to a circularity in the data generation process leading to drug targets much more likely being hub genes (mean log-degree=4.1 vs 3.1, P_{diff}=1.7e-242). In conclusion, we systematically assessed strategies to prioritize drug target genes (using GWAS, molecular QTLs and network information) highlighting promises and potential pitfalls of current approaches.

S69. Methods and databases: Open, benchmarked and FAIR

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 455. The Association to Function Knowledge Portal: An open-access resource for translating variant associations to biological knowledge

Authors:

M. Brandes[1], M. C. Costanzo[1], M. Duby[1], Q. Hoang[1], D-K. Jang[1], R. Koesterer[1], P. Kudtarkar[2], A. Moriondo[1], T. Nguyen[1], O. Ruebenacker[1], P. Smadbeck[1], Y. Sun[2], K. Gaulton[2], M. Boehnke[3], J. Flannick[4], N. Burtt[1]; [1]The Broad Inst., Cambridge, MA, [2]Univ. of California San Diego, San Diego, CA, [3]Univ. of Michigan, Ann Arbor, MI, [4]Boston Children's Hosp., Boston, MA

Abstract:

Most signals from genome-wide association studies (GWAS) result from as-yet-unknown alterations of molecular or cellular function. The knowledge gaps surrounding the functional effects of GWAS associations inhibit the understanding and treatment of human disease. We have created the Association to Function Knowledge Portal (A2FKP; a2fkp.org), the first resource of its kind, to address this gap. The A2FKP aggregates and integrates multiple data types for 435 traits across 11 common disease areas: full GWAS summary statistics from nearly 500 datasets, many of which are not available elsewhere; tissue-specific epigenomic annotations from over 4000 datasets, via the Common Metabolic Disease Genome Atlas (CMDGA; cmdga.org); manually curated credible sets; and over 3 expert-generated effector gene lists. Bioinformatic methods are implemented at scale across these data: novel sample overlap-aware genetic association meta/analysis meta-analysis; predictions of variant effects (VEP and BASSETT), gene-level association analysis (MAGMA); credible set calculations; and annotation global enrichments (GREGOR and S-LDSC). The integrated results are accessible via an open-access web portal, which offers both forward and reverse genetics workflows, with results publicly available through direct download and programmatic APIs. In a forward genetics approach, researchers can query a trait or disease and see genome-wide variant- and gene-level genetic associations and tissue-specific epigenomic annotations that are globally enriched for those genetic associations. In a reverse genetics approach, the Region, Gene, and Variant pages distill and summarize results for a gene or variant of interest. Interactive tools allow researchers to perform custom, on-the-fly analyses, such as generating gene-level association scores based on protected individual-level data and assessing the weight of genetic evidence linking a gene to a disease. The Variant Sifter visualizer allows users to explore, filter, and prioritize genetically associated variants, credible sets, and tissue-specific epigenomic annotations across a region, providing decision support for the prioritization of variants and genes for further research. We work directly with disease communities to provide their expertise along with contributing data or methods within a common knowledge base. In addition, we have built over 8 community portals tailored to each disease community to further accelerate their research.

S70. Molecular investigations into disease mechanisms

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 458. Long-read sequencing and profiling of RNA-binding proteins reveals the pathogenic mechanism of aberrant splicing of an *SCN1A* poison exon in individuals with epilepsy

Authors:

H. C. Happ[1], P. N. Schneider[2], E. Goes[1], J. H. Hong[1], M. Bandouil[1], C. G. Biar[1], K. Engel[3], I. E. Scheffer[4], H. C. Mefford[5], J. D. Calhoun[1], G. L. Carvill[1]; [1]Northwestern Univ., Chicago, IL, [2]Louisiana State Univ., Baton Rouge, LA, [3]Univ. of Colorado Denver, Aurora, CO, [4]The Univ. of Melbourne, Austin Hlth., Heidelberg, Australia, [5]St. Jude Children's Res. Hosp., Memphis, TN

Abstract:

Pathogenic variants in *SCN1A* cause Dravet syndrome (DS), which is a severe developmental and epileptic encephalopathy (DEE). Individuals with DS typically present with febrile seizures around six months, evolve to have multiple refractory seizure types, and are at high risk for sudden death in epilepsy. Approximately 90% of individuals with DS have a *de novo* *SCN1A* truncation or loss-of-function missense variant. We previously identified five variants in five individuals with DS or related *SCN1A* phenotypes that fall in or near a poison exon in *SCN1A* intron 20. We hypothesized these variants lead to increased inclusion of the poison exon, which introduces a premature stop codon, and, therefore, reduced abundance of the full-length *SCN1A* transcript and $Na_v1.1$ protein; this results in *SCN1A* haploinsufficiency and the resulting phenotype. For three of the variants flanking the poison exon, we differentiated patient-derived induced pluripotent stem cells (iPSCs) into neurons. Using long-read sequencing in these patient-specific neurons, we defined the *SCN1A* transcript landscape and revealed how the poison exon is spliced into the *SCN1A* isoform. Patient-specific neurons also showed reduced abundance of $Na_v1.1$ protein as detected by Western blot. To interrogate whether these variants alter any RNA-binding protein (RBP) consensus sites that could account for the aberrant splicing of the poison exon, we used a splicing reporter construct containing the *SCN1A* genomic region with the poison exon and the constitutive exons directly upstream and downstream. We transfected individual constructs containing the wildtype (WT) and patient variants into HEK293T cells and performed RNA-antisense purification with mass spectrometry to identify RBPs. We identified several RBPs that differentially interact with constructs containing the variants as compared to WT constructs, including SRSF1, SRSF3, and HNRNPR. To validate these interactions, we performed RNA immunoprecipitation followed by droplet-digital PCR for the canonical and poison exon-inclusion *SCN1A* transcripts. Overall, we confirm aberrant *SCN1A* poison exon splicing and $Na_v1.1$ haploinsufficiency in patient neurons and that this altered splicing is facilitated by the RBPs SRSF1/3 and HNRNPR. By establishing the pathogenic mechanism by which genetic variants can alter poison exon splicing through the disruption of RBP interactions, this work lays the foundation for developing a framework to identify poison exons via RBP consensus motifs and determine which variants in individuals with neurodevelopmental disorders may disrupt poison exon splicing.

S70. Molecular investigations into disease mechanisms

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 459. 5,510 bp deletion upstream of NCK2 is associated with Alzheimer's disease and deletes 6 CREs in microglia PU.1 super-enhancer

Authors:

J. Farrell[1], C. Zhu[1], X. Sun[1], J. Huang[1], C. Dalgard[2], B. Dombroski[3], J. Malamon[4], Y. Leung[3], E. R. Martin[5], W. S. Bush[6], M. Pericak-Vance[5], B. Vardarajan[7], W-P. Lee[3], J. Haines[6], G. Schellenberg[3], L. Farrer[1]; [1]Boston Univ. Sch. of Med., Boston, MA, [2]Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [3]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [4]Univ. of Colorado Sch. of Med., Aurora, CO, [5]Univ. of Miami Miller Sch. of Med., Miami, FL, [6]Case Western Reserve Univ., Cleveland, OH, [7]Columbia Univ., New York, NY

Abstract:

Purpose
While the TOPMed imputation panel is an excellent resource to impute SNPs into GWAS datasets, the panel does not include structural variants (SVs). To test the association of SVs with late onset Alzheimer disease (AD), we developed an imputation panel containing 170,073 SVs from the Alzheimer Disease Sequencing Project (ADSP) R1 Cohort (N=4,789). We imputed these SVs into large GWAS datasets to test for association with AD and proxy-AD outcomes.
Methods
We compiled a library of SV breakpoints discovered with Scalpel, Manta and Parliament pipelines in the ADSP Cohort which were joint-genotyped using Graphtyper. We imputed the SVs into the UK Biobank (N=318,651 British) and Alzheimer Disease Genetics Consortium (ADGC, N= 27,244 Caucasian) GWAS datasets using Minimac4, and tested their association with AD status using a logistic regression model with co-variates for age, gender, and 10 PCs for population substructure.
Results
We identified genome-wide significant association (OR=1.55, $P=6.82x10^{-9}$) of AD with a rare 5,510 bp deletion (MAF=0.005) located 12,735 to 18,245 bp upstream of NCK2. This deletion knocks out 6 ENCODE candidate Cis Regulatory Elements (cCRE) located in a microglia PU.1 super-enhancer. The transcription factor PU.1 is a master regulator of myeloid cells and controls microglial development and function. Chromatin H3K27ac activity is enriched at this PU.1 super-enhancer site in microglia but not neurons, oligodendrocytes and astrocytes where activity is limited to the NCK2 promoter. The Human Protein Atlas scRNA experiments show that microglia (77.8 nTPM) have higher expression than neurons (Excitatory: 60.0 nTPM, Inhibitory:48.4 nTPM), oligodendrocytes (18.3 nTPM) and astrocytes (28.1 nTPM) as expected with the additional active PU.1 enhancer. This is also consistent with previous results that show a majority of AD associated genes are exclusively or most highly expressed in microglia. With Sanger sequencing in 132 ADSP samples, we found the deletion to be in complete linkage disequilibrium with the previously detected AD-associated NCK2 intronic variant (rs143080277).
Conclusion
The imputation of structural variants is a useful approach for discovering novel disease associations in large GWAS datasets. The location of the associated intergenic SV suggests that the deletion is functional and could decrease NCK2 expression in microglia given its removal of 6 cCREs in a PU.1 Super-enhancer. NCK2 plays a role in ELK1-dependent transcriptional activation, which may be involved in the formation of long-term memory. Overall, these results provide further evidence that both NCK2 and microglia are associated with AD risk.

S70. Molecular investigations into disease mechanisms

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 460. Genomic discovery and functional validation of MRP1 as a novel therapeutic target for sickle cell disease

Authors:

S. Lessard[1], Y. Hara[1], E. Kawabata[2], V. T. Lemgart[1], P. G. Bronson[3], A. Hicks[1], R. Peters[4], S. Krishnamoorthy[4], N. Watkins[5], D. J. Roberts[6], J. Eglinton[7], E. Di Angelantonio[2], J. Danesh[2], W. Astle[2], D. Paul[2], A. Butterworth[2]; [1]Sanofi, Cambridge, MA, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Biogen, Cambridge, MA, [4]Sanofi, Waltham, MA, [5]Natl. Hlth.Service Blood and Transplant, Cambridge, United Kingdom, [6]Natl. Inst. for Hlth.Res. Oxford BioMed. Res. Ctr., Oxford, United Kingdom, [7]Oxford Univ. Hosp. NHS Fndn. Trust, Oxford, United Kingdom

Abstract:

Induction of fetal hemoglobin (HbF) is a viable therapeutic strategy for treating sickle cell disease (SCD). Genome-wide association studies (GWAS) identified loci associated with HbF regulation, including variants modulating the transcription factor *BCL11A,* a target in gene therapy for SCD. However, access to this therapeutic approach remains a challenge. In addition, hydroxyurea, a HbF inducer and standard of care in SCD, has limited effectiveness and carries some concerns for carcinogenicity associated effects. Therefore, there is a large unmet need for novel accessible therapies for SCD. To identify potential therapeutic targets for SCD, we performed a GWAS for HbF expression in 11,004 healthy participants from the INTERVAL study. We identified 11 genome-wide significant loci, 8 of which were novel. We prioritized putative causal variants by fine-mapping, including a low-frequency missense variant in *ABCC1* suggesting that this gene, which encodes the efflux pump for glutathione Multidrug Resistant Protein 1 (MRP1), is a potential modulator of γ-globin expression (rs60782127, p.R433S, β=0.4SD, $P<5x10^{-14}$, posterior probability=0.99). This variant was strongly associated with plasma glutathione levels in the INTERVAL study (β=0.8SD, $P<3x10^{-38}$). To elucidate the role of *ABCC1* in HbF regulation, we performed a CRISPR-Cas9-mediated *ABCC1* knockdown in erythroid HUDEP2 cells and CD34+ hematopoietic stem and progenitor cells from healthy donors. In CD34+ cells, *ABCC1* knockdown (84%) increased HbF (2-3-fold, $P<0.05$) and intracellular glutathione level (6-8-fold, $P<0.01$) after differentiation. Similarly, HUDEP2 clones homozygous for the *ABCC1*-R433S allele showed increases in HbF (6-fold, $P<0.01$) and glutathione (17-fold, $P<0.01$). Pharmacological inhibition of MRP1 in CD34+ cells by the selective inhibitor MK571 led to an increase in HbF positive cells (2-fold, $P<0.05$) and glutathione (5-fold, $P<0.01$), and prevented sickling of CD34+ cells from SCD patients in hypoxic condition ($P<0.05$), without impacting erythroid differentiation. Moreover, transcriptomic, proteomic, and phospho-proteomic analyses revealed an activation of the NRF2-mediated oxidative stress pathway in MK571-treated cells. Addition of NL385, an NRF2 inhibitor, in MK571-treated CD34+ cells abolished HbF induction, confirming a NRF2-dependent HbF induction. Overall, we identified 8 novel loci associated with HbF, including *ABCC1*/MRP1. We validated MRP1 as a regulator of HbF through NRF2-mediated pathway, and showed that MRP1 inhibition prevents sickling of CD34+ cells from SCD patients, establishing MRP1 as a potential new therapeutic target in SCD.

S70. Molecular investigations into disease mechanisms

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 461. DNA sequence is the primary determinant of R-loop formation across genomes

Authors:

L. Walker, A. Stanely, S. Hartono, F. Chedin; Univ. of California, Davis, Davis, CA

Abstract:

R-loops are hybrid structures of DNA and nascent RNA that form widely across eukaryotic and bacterial genomes during transcription and can lead to transcriptional repression and genome instability. Factors such as primary DNA sequence, nascent RNA levels, DNA and RNA binding factors, and chromatin state are thought to influence R-loop metabolism. What is unknown are the relative contributions of these factors towards R-loop formation, and to what extent the underlying rules governing R-loop formation are conserved across species. To address this, we constructed a deep neural network to predict R-loop formation along the genome that is based on primary DNA sequence and other genomic and epigenomic annotations. Our genomic annotations included those related to nascent and mature transcription, chromatin state (histone marks, methylation), and chromatin accessibility. Surprisingly and despite R-loop formation's dependence on transcription, we found that DNA sequence was the most predictive factor for R-loop formation and frequency; performance of the DNA sequence-only model was 41% higher than the nascent transcription-only model. When using a DNA sequence-only model as a baseline, we found that adding epigenetic state and nascent transcription data provided modest additional prediction power (5% more). This suggests that the role of DNA sequence is distinct from its influence on transcription and epigenetic state. When interrogating the neural network to identify the most salient DNA sequence features that predicted R-loop formation, we found that local islands of G's and T's provided strikingly high contributions (positively and negatively, respectively). Simulations removing those islands yielded substantial decrease in predicted R-loop formation. Finally, an open question is the extent to which the underlying molecular rules governing R-loop formation is conserved across species. To probe this, we trained individual models on human, mouse, zebrafish, chicken, fruit fly, and roundworm datasets and evaluated how well a model trained on one species could predict R-loop data of another species. We found that models trained in human were 84% as predictive of yeast R-loop formation as other models trained directly on yeast, despite the large variance in GC content and gene density between their genomes. Our results suggest high conservation of sequence rules governing R-loop formation.

S70. Molecular investigations into disease mechanisms

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 462. Chromatin accessibility is a potential factor contributing to the lower risk effect of the *APOE* ε4 allele in individuals of African ancestry

Authors:

J. Young[1], K. Celis[1], M. Muniz Moreno[2], F. Rajabli[1], P. Whitehead[1], K. Hamilton-Nelson[3], D. Dykxhoorn[1], K. Nuytemans[2], L. Wang[1], C. Dalgard[4], M. Flanagan[5], S. Weintraub[5], C. Geula[5], M. Gearing[6], D. Bennett[7], F. Jin[8], T. Schuck[9], M. Pericak-Vance[3], A. Griswold[1], J. Vance[1]; [1]Univ. of Miami, Miami, FL, [2]John P. Hussman Inst. for Human Genomics, Miami, FL, [3]Univ. of Miami Miller Sch. of Med., Miami, FL, [4]USUHS - TAGC, Bethesda, MD, [5]Northwestern Univ., Chicago, IL, [6]Emory Univ., Atlanta, GA, [7]Rush Univ. Med. Ctr., Chicago, IL, [8]Ludwig Inst. for Cancer Res., San Diego, CA, [9]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Introduction: The ε4 allele at the APOE gene confers different risks for Alzheimer's disease (AD) across ancestry groups. Individuals of African ancestry have a lower *APOE* ε4-related AD risk compared with European. The ancestry-related risk is driven by the genomic region surrounding APOE (termed Local Ancestry; LA). We recently showed, using single nucleus RNA sequencing (snRNA-seq) data, that AD patients with European LA (ELA) expressed significantly greater levels of *APOE* than African LA (ALA) *APOE* ε4 carriers, suggesting that this locus is differentially regulated between ancestries. We explored the hypothesis that the *APOE* expression difference was due to chromatin accessibility differences in ELA and ALA carriers. **Methods:** We performed single nuclei Assays for Transposase Accessible Chromatin sequencing (snATAC-seq) coupled with snRNA-seq from frozen frontal cortex of six ALA and six ELA AD patients homozygous for *APOE* ε4. snATAC-seq data was processed using the ArchR package and snRNA-seq using Seurat for data integration, and differential analysis. We performed analysis of transcription factor binding analysis using MEME. **Results:** We determined that the region surrounding *APOE,* including its promoter area*,* have greater chromatin accessibility in ELA than ALA carriers in astrocytes, corresponding to the greater *APOE* ε4 expression in ELA carriers. We inferred transcription factors that could potentially regulate *APOE* expression in astrocytes from the analysis of chromatin accessibility, gene expression and DNA binding data in the differentially accessible peaks in the *APOE* promoter. The increased chromatin accessibility in European ancestry astrocytes extended beyond the LA region and was observed genome-wide. Genes with differential accessibility and expression in ELA vs. ALA astrocytes were enriched for synaptic function, cholesterol processing and astrocyte reactivity. **Conclusion:** Our results suggest that chromatin accessibility differences contribute to the differential expression of *APOE* in ELA vs. ALA *APOE* ε4 carriers, corresponding to the increased AD risk in ELA. The data supports that the reduction of *APOE4* expression, especially in astrocytes, is a potential therapeutic approach for Alzheimer Disease.

S70. Molecular investigations into disease mechanisms

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 463. Bi-allelic variants in SPOUT1, an RNA methyltransferase functioning in spindle organization, cause a novel neurodevelopment disorder

Authors:

A. Dharmadhikari[1], M. Abad[2], F. Ullah[3], T. Sands[4], D. Marafi[5], T. Mitani[5], S. Khan[6], K. Moore[7], D. Koboldt[8], T. Mosher[9], S. Hickey[10], D. Albert[9], C. Umandap[11], S. Berger[12], N. Lippa[13], B. Popp[14], S. Rentas[15], A. Ritter[16], K. Izumi[16], L. Conlin[17], H. Norwood[18], A. Lewanda[18], H. Gabriel[19], D. Pehlivan[5], D. Goldstein[20], A. Iglesias[21], R. Jamra[22], J. Rosenfeld[23], J. Posey[5], I. Samejima[2], J. Lupski[5], V. Aggarwal[24], S. Ohta[25], W. Earnshaw[2], E. Davis[26], A. Jeyaprakash[27], J. Liao[12]; [1]Children's Hosp. Los Angeles, Los Angeles, CA, [2]Wellcome Ctr. for Cell Biology, Univ. of Edinburgh, Edinburgh, United Kingdom, [3]Lurie children s Hosp. of Chicago, chicago, IL, [4]Div. of Child Neurology & Inst. for Genomic Med., Columbia Univ. Irving Med. Ctr.; New York Presbyterian Hosp., New York, NY, [5]Baylor Coll. of Med., Houston, TX, [6]Ann & Robert H. Lurie children's Hosp. of Chicago, Chicago, IL, [7]Stanley Manne Children's Res. Inst., Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, [8]Steve and Cindy Rasmussen Inst. for Genomic Med. at Nationwide Children's Hosp.; Dept. of Pediatrics, The Ohio State Univ. Coll. of Med., Columbus, OH, [9]Nationwide Children's Hosp., Columbus, OH, [10]Dept. of Pediatrics, The Ohio State Univ. Coll. of Med.; Div. of Genetic & Genomic Med., Nationwide Children's Hosp., Columbus, OH, [11]Childrens Rehabilitative Services Clinic, Phoenix, AZ, [12]Columbia Univ. Med. Ctr., New York, NY, [13]Inst. for Genomic Med., Columbia Univ., New York City, NY, [14]Univ. of Leipzig Hosp. and Clinics, Leipzig, Germany, Leipzig, Germany, [15]Div. of Molecular Pathology, Genetics and Genomics, Dept. of Pathology, Duke Univ., Durham, NC, [16]Children's Hosp. of Philadelphia, Philadelphia, PA, [17]Children s Hosp. of Philadelphia, Philadelphia, PA, [18]Children's Natl. Rare Disease Inst., Washington, DC, [19]Praxisfür Humangenetik Tübingen, Tübingen, Germany, [20]Columbia Univ Med Ctr, New York, NY, [21]Columbia Univ Med Ctr, HASTINGS ON HUDSON, NY, [22]Univ. Med. Ctr., Leipzig, Germany, [23]Dept. of Molecular and Human Genetics, Baylor Coll. of Med.; Baylor Genetics Lab., Houston, TX, [24]Columbia Univ, New York, NY, [25]Inst. for Genetic Med. Pathophysiology, Hokkaido Univ., Sapporo, Japan, [26]Lurie Children s Hosp. of Chicago, Chicago, IL, [27]Children's Hosp. Los Angeles Wellcome Ctr. for Cell Biology, Univ. of Edinburgh, Edinburgh, United Kingdom

Abstract:

Faithful chromosome segregation during mitosis depends on the formation of a functional mitotic spindle. SPOUT1, an essential putative SPOUT RNA methyltransferase, was identified in a proteomics screen as a novel mitotic chromosome associated protein. SPOUT1 depletion in human cell lines leads to centrosome detachment from the spindle poles and chromosome misalignment, while homozygous Spout1 knockout mice show early embryonic lethality before implantation, suggesting its importance in spindle organization and embryonic development. However, its role in human normal development and rare disease traits remains unknown.

Using a combination of exome sequencing (ES) and whole genome homozygosity mapping we have identified twelve individuals from nine unrelated families from across the globe with bi-allelic variants in *SPOUT1* via a multi-institutional collaboration aided by GeneMatcher. Common phenotypes observed include severe to profound intellectual disability and/or developmental delays (ID/DD) (12/12; 100%), short stature (6/7; 86%), failure to thrive (5/6; 84%). epilepsy (9/11; 81%), axial hypotonia (9/11; 81%), and microcephaly (6/8; 75%). In a majority of subjects (7/9), the phenotype was consistent with epileptic encephalopathy (EE) with history of infantile spasms/hypsarrhythmia in 5/7, Lennox-Gastaut syndrome in 1/5, and unclassified EE in 1/5.

The *SPOUT1* variant alleles detected in these patients include ten missense variants, one frameshift duplication and one in-frame deletion.

To understand the structural basis for SPOUT1 function we determined its high-resolution crystal structure. Structural characterization shows that SPOUT1 forms a homodimer via its catalytic domain. Functional studies of three SPOUT1 missense variants seen in more than one family show reduction in methyl transferase activity and compromise SPOUT1 tethering of centrosomes to the spindle poles at varying degrees. siRNA based functional rescue assays using structure-guided mutants show that the methyltransferase activity of SPOUT1 is essential for its function. Additional studies in zebrafish show that spout1 ablation leads to a significant reduction of larval head size.

In the literature, several similar genes involved in chromosome alignment and/or encoding for an RNA methyltransferase have also been reported to be associated with neurodevelopmental disorders (NDDs), including *NSUN2, MCPH1, CHAMP1,* and *CASC5*. In summary, we propose *SPOUT1* as a novel autosomal recessive NDD gene and provide structural and functional evidence for its role in mitotic spindle organization and neural development.

S71. Preconception carrier screening and rapid WGS of newborns

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 466. Outcomes of reproductive genetic carrier screening in 9,111 couples screened for 1,281 autosomal and X-linked genes: results from the Australian Reproductive Genetic Carrier Screening Project ("Mackenzie's Mission")

Authors:

E. Kirk[1], A. Archibald[2], S. Best[3], T. Boughtwood[3], J. Caruana[3], M. Downes[4], A. Newson[5], P. Scuffham[6], N. Laing[7], M. Delatycki[2], Mackenzie's Mission Research Team; [1]New South Wales Hlth.Pathology, Sydney, Australia, [2]Victorian Clinical Genetics Service, Melbourne, Australia, [3]Australian Genomics Hlth.Alliance, Melbourne, Australia, [4]Ctr. for Applied Hlth.Economics, Sch. of Med. and Dentistry, Griffith Univ., Gold Coast, Australia, [5]Sydney Hlth.Ethics, Sydney Sch. of Publ. Hlth., Faculty of Med. and Hlth., Univ. of Sydney, Sydney, Australia, [6]Menzies Hlth.Inst. Queensland, Gold Coast, Australia, [7]Harry Perkins Inst. of Med. Res., Perth, Australia

Abstract:

Background: The Australian Reproductive Genetic Carrier Screening Project (Mackenzie's Mission), is a government funded research project which aims to investigate all aspects of reproductive genetic carrier screening (RGCS) to guide future population wide implementation.
Methods: Participants were recruited by participating healthcare practitioners (HCPs) during routine medical care. Recruitment involved discussion with the HCP followed by interactive online education, consent, and completion of research surveys. Participants were mailed mouth swabs for home sample collection. Three laboratories provided screening for 1,281 genes. Analysis was on a couples basis, with individual results reported only for X-linked conditions. Variants were reported only if classified as Likely Pathogenic or Pathogenic. Survey data were collected at enrolment, at testing and 3 and 12 months post-result.
Results: 761 participating HCPs recruited 9,111 couples who completed enrolment, agreed to screening and returned samples. Couples were broadly representative of Australian reproductive-age population geographic distribution and socio-economic status. Pre-test information and consent was successfully delivered via an online portal with high knowledge post-education (17,438 participants (96%) scoring ≥70%) and low decisional conflict (n=7804, mean=10.5, SD=11.8, where <25/100 indicates low decisional conflict) and regret (n=4189, median=0, IQR=0-15, where 0/100=no regret). For 8,564 couples with results available at abstract submission, 2.1% received a new high-risk result. Of these, 80% were for genes other than *CFTR*, *SMN1* and *FMR1* and 43% involved genes not included in the ACMG Carrier Screening Practice Resource Tier 3 gene lists. By contrast, individual analysis would have led to 90% of individuals receiving one or more carrier results, with 98% of reports having no direct clinical utility for the couple. 68% of high-risk couples took up or intend to take up the offer of a reproductive intervention (19% still deciding). Couples receiving a low-risk result did not experience increased anxiety on STAI scores (n=3961, mean=32.8, SD=11.2, where ≥40 indicates clinically meaningful anxiety). Economic modelling showed that RGCS would be highly cost-effective. A comprehensive analysis of ethical issues was undertaken.
Conclusions: Couple based RGCS for a very large panel of genes has a high yield, is cost-effective, and can be successfully delivered via an online platform.

S71. Preconception carrier screening and rapid WGS of newborns

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 467. Carrier rates for prenatal-lethal genomic variants: A case for expanded genetic screening

Authors:

M. Aminbeidokhti[1], J-H. Qu[1], B. Monk[1], S. Belur[1], S. Yatsenko[2], A. Rajkovic[1]; [1]Univ. of California San Francisco, San Francisco, CA, [2]Univ Pittsburgh, Magee-Womens Hosp., Pittsburgh, PA

Abstract:

Purpose: Rapid advances in genome sequencing technology have expanded carrier screening and prenatal clinical genetic testing to assess for many severe and potentially lethal inherited conditions at a nominal cost. For prenatal phenotypes, however, expanded carrier screening has a diagnostic yield of ~10%, while causes of a miscarriage of euploid conceptions are mainly unknown. It is likely that hundreds of genes critical for fetal development have not yet been linked to prenatal lethal phenotypes in humans. This study used a bioinformatic approach to estimate population carrier rates for known and novel recessive lethal Mendelian conditions.
Methods: A list of lethal genes associated with recessive lethal conditions has been derived from the Online Mendelian Inheritance in Man database. Pathogenic, likely pathogenic, and loss-of-function variants (n=16,603) for the lethal genes were downloaded from the Genome Aggregation Database (gnomAD v2.1.1) and ClinVar and used to calculate a theoretical risk of recessive lethal genotypes spanning seven ethnic groups (African/African American [afr], Hispanic [amr], Ashkenazi Jewish [asj], East Asian [eas], Finnish [fin], Non-Finnish European [nfe], South Asian [sas]) and subgroups of unrelated individuals. Population risk probabilities for the inheritance of lethal genotypes were estimated using variant carrier rate (VCR), gene carrier rate (GCR), cumulative carrier rate (CCR), and at-risk couple rate (ACR) metrics.
Results: The study found 446 human lethal genes with at least one variant meeting the search criteria in gnomAD. Although the VCR<0.0001 and GCR<0.05 were very low across populations, the accumulative CCRs (0.3-0.7) and ACRs (0.03, except afr) were high. Between 41.0% (amr) and 66.5% (afr) of individuals were found to carry a variant in at least one lethal gene. Analysis of intra-ancestry couples showed the highest ACR amongst afr/afr (0.09) and asj/asj (0.03) couples. Analysis of inter-ancestry couples indicated that asj/nfe (0.02) couples were also at relatively high risk. Overall, screening all 446 lethal genes identified 0.18-8.63% of couples as being at risk for having a conception or a child affected by a lethal condition.
Conclusion: This study identified and ranked a set of genes associated with adverse prenatal and perinatal outcomes that may present in individuals across ethnic backgrounds. The diversity of these genes amongst the various ethnic groups underscores the importance of designing more individualized carrier screening panels for prospective parents while they are navigating their reproductive options.

S71. Preconception carrier screening and rapid WGS of newborns

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 468. The PREGCARE study: Personalized recurrence risk assessment following the birth of a child with a pathogenic *de novo* mutation

Authors:

A. Goriely[1], M. Bernkopf[1], U. B. Abdullah[1], S. J. Bush[1], K. Wood[1], S. Ghaffari[1], E. Giannoulatou[2], N. Koelling[1], G. Maher[1], L. M. Thibault[3], J. Williams[4], E. Burkitt-Wright[5], A. Bloss[4], A. Dixit[6], J. Eason[6], F. Elmslie[7], U. Kini[4], E. Kivuva[8], M. Lees[9], A. H. Nemeth[4], K. Saunders[4], D. J. Shears[4], M. Splitt[10], H. Stewart[4], P. Clouston[4], PREGCARE Clinical Collaborative Group, R. W. Davies[1], A. O. M. Wilkie[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Victor Chang Cardiac Res. Inst., Darlinghurst, Australia, [3]Garvan Inst. of Med. Res., UNSW Sydney, Australia, [4]Oxford Univ. Hosp. NHS Fndn. Trust, Oxford, United Kingdom, [5]Manchester Univ. NHS Fndn. Trust, Manchester, United Kingdom, [6]Nottingham Univ. Hosp. NHS Trust, Nottingham, United Kingdom, [7]St George's Univ. Hosp. NHS Fndn. Trust, London, United Kingdom, [8]Royal Devon & Exeter NHS Fndn. Trust, Exeter, United Kingdom, [9]Great Ormond Street Hosp. NHS Fndn. Trust, London, United Kingdom, [10]The Newcastle upon Tyne Hosp. NHS Fndn. Trust, Newcastle, United Kingdom

Abstract:

Following the birth of a child with a disorder caused by a *de novo* mutation (DNM), couples are typically given a risk of 1-2% for recurrence in a subsequent pregnancy. But this figure represents a population average that is nearly always incorrect for the specific couple. In most, the risk is negligible, whereas in a minority (those in which one of the parents is a germline mosaic) the risk might reach 50%.

The UK-based PREGCARE (PREcision Genetic Counselling And REproduction) study has developed a systematic approach to providing personalized recurrence risk stratification. We recruited 58 families (total of 59 DNMs in 49 different genes) who had a child with a serious developmental disorder caused by an apparent DNM and were seeking individualized reproductive counselling about recurrence risk in a future pregnancy.

The PREGCARE strategy relies on the collection of up to 14 tissues (including paternal semen) from the mother-father-proband trio that were analysed through two complementary methods: (1) Deep-sequencing (>5000x, Illumina) of the DNM identified 6 couples (10%) in whom one of the parents was mosaic and one case (1.7%) of post-zygotic mosaicism in the proband. This approach allowed us to efficiently single out the vast majority of couples at greatest risk for recurrence. Moreover, sperm DNA analysis provided a direct risk quantification for the 4 paternal cases. (2) For the 52 remaining DNMs, long-read Nanopore sequencing was used to determine the parent-of-origin of the DNM via haplotyping, which further reassured a majority of couples (65%) that the DNM had likely occurred as a paternal one-off event during spermatogenesis and that their recurrence risk was very low (<0.1%). For maternally-derived DNMs or for those couples in whom the haplotype could not be resolved, we estimated the risk to be reduced by ~2- to 10-fold compared to the population baseline, respectively.

Our data demonstrate that for all couples, it is possible to refine the recurrence risk and in the majority of cases the risk of having another affected child is in fact very low, potentially reducing anxiety and the need for expensive pre-implantation or prenatal diagnostic options. For couples in whom we detected overt mosaicism, the risk is higher (and quantifiable through sperm analysis for the paternal cases). Providing evidence-based estimation of the risk will allow these couples to make informed choices about the different diagnostic options available to them. Overall, we show that providing pre-conception recurrence risk assessment to couples can be achieved and offers the prospect of driving a major transformation in the practice of genetic counselling.

S71. Preconception carrier screening and rapid WGS of newborns

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 469. A genome sequencing system for universal newborn screening, diagnosis, and precision medicine for severe genetic diseases

Authors:

W. Mowrey[1], S. F. Kingsmore[2,3,4], L. D. Smith[2], C. M. Kunard[5], G. del Angel[1], D. P. Dimmock[2,3], Y. Ding[2,3], A. Feigenbaum[2,3,6], E. Frise[7], R. C. Green[8], K. Hall[5], C. Hansen[2,3], C. Hobbs[2,3], M. Kiel[9], Y. H. Kwon[2,3], L. Madhavrao[2,3], S. Lefebvre[1], R. Mardach[2,3,6], D. Oh[2,3], G. Powley[10], G. Scharer[2], S. Shelnutt[10], S. S. Mehtalia[5], S. Papadopoulos[10], S. Schwartz[9], M. G. Reese[7], M. Wright[2,3], M. J. Willis[2], A. R. Wolen[10], T. Defay[1]; [1]Alexion AstraZeneca Rare Disease, Boston, MA, [2]Rady Children's Inst. for Genomic Med., San Diego, CA, [3]Rady Children's Hosp., San Diego, CA, [4]Keck Graduate Inst., Claremont, CA, [5]Illumina, Inc., San Diego, CA, [6]Dept. of Pediatrics, Univ. of California San Diego, San Diego, CA, [7]Fabric Genomics, Inc., Oakland, CA, [8]Mass Gen. Brigham, Broad Inst., Ariadne Labs and Harvard Med. Sch., Boston, MA, [9]Genomenon, Inc., Ann Arbor, MI, [10]TileDB, Inc., Cambridge, MA

Abstract:

Newborn screening (NBS) dramatically improves outcomes in severe, childhood genetic disorders by treatment at or before symptom onset. In many genetic diseases with effective treatments, however, outcomes remain poor since NBS has lagged behind orphan drug development. Rapid whole genome sequencing (rWGS) is attractive for comprehensive NBS since it concomitantly examines almost all genetic diseases and is gaining acceptance as a first-tier test for genetic disease diagnosis in critically ill newborns. We describe prototypic methods for scalable, parentally consented, feedback-informed NBS and diagnosis of genetic diseases by rWGS and virtual, acute management guidance (NBS-rWGS). NBS-rWGS data also resides in a secure, parent-controlled data platform with transparent rights that convey to the child, persist across the lifetime, and can be integrated into the child's future medical care. Using established criteria and a modified Delphi technique, we reviewed 457 genetic diseases, and retained 388 (85%) as suitable for NBS-rWGS. Simulated NBS-rWGS for 388 disorders in 454,707 exome sequences of UK Biobank subjects with 29,865 pathogenic or likely pathogenic variants associated with these diseases had specificity of 99.7% following root cause analysis. In 4,376 critically ill children with suspected genetic disorders and their parents, simulated NBS-rWGS for 388 disorders identified 109 (87%) of 125 findings previously made by diagnostic rWGS, and 23 findings not previously reported. The negative predictive value and sensitivity of NBS-rWGS (99.6% and 89.2%) were a little higher than diagnostic rWGS (99.5% and 84.5%, respectively). In 43 critically ill children previously diagnosed by rWGS and identified by NBS-rWGS, earlier interventions may have avoided severe morbidity and hospitalization. We invite interested groups worldwide to help further refine the set of NBS-rWGS conditions and examine clinical utility and cost effectiveness in prospective clinical studies.
This research has been conducted using the UK Biobank Resource under application numbers 82213 and 26041.

S71. Preconception carrier screening and rapid WGS of newborns

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 470. Experiences of rapid whole genome sequencing (rWGS) at Baylor Genetics: a powerful and comprehensive first-tier genetic test

Authors:

H. Dai[1,2], L. Vossaert[1,2], X. Zhao[1], K. Schulze[1], P. Liu[1,2], C. Qu[2], V. Nguyen[2], M. Santana[2], F. Xia[1,2], C. Eng[1,2], L. Meng[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX

Abstract:

Introduction: Whole genome sequencing with rapid processing and analysis (rWGS) can provide the most comprehensive "all in one" solution by interrogating not only small nucleotide variations (SNV), but also copy number variations (CNV), short tandem repeats (STR), and mitochondrial genome variations (mtDNA). Thus, rWGS can meet the requests from clinicians looking for the best diagnostic tools in challenging situations where diagnostic yield and turnaround time (TAT) are most important for the clinical management of critically ill children with a suspected underlying genetic etiology. Here we summarize our experiences with rWGS in a clinical setting at Baylor Genetics to provide insights for clinicians who require results quickly to guide medical management. **Methods:** We evaluated a total of 150 cases sent for rWGS, including 123 trio cases, 16 duo cases and 11 proband only cases. Demographic data, clinical history, prior testing, TAT and diagnostic findings were assessed. A clinical intensive care setting was considered indicative of critical illness. **Results:** The median age of the 150 tested individuals was 1.8 months, with 40 % (n=60) of them less than one month old. Abnormalities in multiple systems were reported as the largest group of these individuals (40%, n=60), followed by anomalies in the neuro system (26.6%, n=40). Cardiac disorders (11.3%, n=17) was the third most common indication. Nearly half of the patients (46.6%, n=70) were hospitalized in intensive care units. More than two thirds of cases (71.3%, n=102) had no previous genetic testing. The shortest TAT was 3 days, with an average TAT of 4.5 days and >90% of cases reported within 5 days, thanks to recent improvements in both wet lab and dry lab processes. The power of rWGS was illustrated by the fact that one in four solved cases involved mutation types other than SNV, including chromosome abnormalities (n=4), CNV only (n=4), SNV plus CNV (n=3) and STR (n=2). Four cases were identified with a dual diagnosis (7.7%, n=4): one case with a pathogenic variant in mitochondrial variant in addition to pathogenic variants in the nuclear genome, and three cases with a combination of SNV and CNV calling. **Conclusion:** The comprehensive detection power of whole genome sequencing as described above, in the setting of rapid processing and reporting, should be available to clinicians as a first-tier genetic test, especially for critically ill children in intensive care.

S71. Preconception carrier screening and rapid WGS of newborns

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 471. SeqFirst: Improving equitable access for precise genetic diagnosis in critically ill infants

Authors:

T. Wenger[1], M. Sikes[2], A. Scott[2], C. Davis[1], K. Buckingham[1], J. X. Chong[1], A. Keefe[1], K. Dipple[1], D. Miller[1], J-H. Yu[1], D. Veenstra[1], D. Copenheaver[1], K. MacDuffie[1], J. Juusola[3], K. Retterer[3], K. McWalter[1], A. Snook[1], P. Kruszka[3], M. Bamshad[1]; [1]Univ. of Washington, Seattle, WA, [2]Seattle Children's Hosp., Seattle, WA, [3]GeneDx, Gaithersburg, MD

Abstract:

SeqFirst is a project to develop and test approaches for equitable access for a precision genetic diagnosis (PrGD) at the initial point of care for infants less than six months of age with a critical illness using simple exclusion criteria (i.e., presentation fully explained by prematurity, trauma, or infection) rather than complex inclusion criteria to identify infants eligible for rapid whole genome sequencing (rWGS). From January 2021 to January 2022, 411 infants were admitted to the neonatal intensive care unit (NICU) at Seattle Children's Hospital, of which 233 met eligibility criteria and 125 were enrolled. rWGS was performed in parallel to conventional clinical care to enable comparison of diagnostic rates, referral patterns and outcomes. Abnormal rWGS results were found in 79/125 (63%) and prompted a clinical genetics consult to determine whether the variant(s) explained (n=36), likely explained (n=23), partially explained (n=5), or did not explain (n=15) clinical findings. Explanatory rWGS results led to a change in management (COM) in 88% of families. Of the 64 whose findings were likely, partially or fully explained by rWGS, 16 (25%) were not suspected by the NICU team to have a genetic condition. In other words, there was no clinical suspicion of a genetic condition in 25% of newborns with diagnostic rWGS, and of these, rWGS resulted in a change of management in 15 (94%). Notably, 10 (63%) of the families of newborns in whom a genetic condition was not considered prior to diagnostic rWGS identified as non-white. Interviews of providers and families are underway to identify potential explanations for this difference. Simplifying workflows to identify newborns who could potentially benefit from rWGS by ascertaining eligibility upon admission to the NICU using broad but precise inclusion criteria, and supporting provider readiness with on-site clinical genetics services resulted in substantially increased access to rWGS. In contrast to the expectation that increasing accessibility to rWGS would result in a low diagnostic rate, the observed diagnostic rate of 51% was as high if not higher than compared to studies prioritizing cases in which there is high suspicion of a genetic condition. The limitations of a small research staff and reliance on face-to-face encounters that are labor and time intensive suggests a need to move toward technologies to increase efficiency, improve scalability, potentially reduce costs, and more importantly improve access to rWGS. Collectively, our results demonstrate clear opportunities exist to improve equitable access to enable precise genetic diagnosis that will benefit critically ill newborns.

S72. Sequencing vs panel testing: Is more better?

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 474. Whole genome sequencing is a powerful “one-stop shop” screening assay for uncovering undiagnosed conditions in apparently healthy pediatric cohort

Authors:

J. Balciuniene, R. Liu, C. Collins, L. Bean, F. Guo, B. Nallamilli, N. Guruju, X. Chen-Deutsch, R. Yousaf, K. Fura, E. Chin, C. da Silva, A. Mathur, Z. Ma, M. Hegde; PerkinElmer Genomics, Pittsburgh, PA

Abstract:

Whole genome sequencing (WGS) is a powerful genomic diagnostics tool. It provides higher clinical sensitivity than exome testing, in part due to a uniform coverage across the genome facilitating copy number variant (CNV) calls, and inclusion of deep intronic variant and mitochondrial DNA (mtDNA) analysis. With decreased sequencing costs, WGS is becoming a feasible option for screening healthy people. Benefits of WGS-based healthy screening include uncovering genetic risks beyond those found in standard carrier and/or newborn screening, informing timely medical management efforts, and identifying variants of pharmacogenomic (PGx) significance. Here we present data from WGS-based healthy screening test for a pediatric population that covers 22000 genes and analyzes 2500 genes relevant for pediatric-onset diseases. Only diagnostic findings (Dx) and PGx variants are reported. Carrier status or diagnoses in adult-onset disorders are not included in pediatric reports. Since 2018, we performed WGS on 532 reportedly asymptomatic pediatric subjects (51% females and 49% males). The median age at testing was 33 days of age, with consent for newborns obtained at pre-delivery counseling. Dried blood spots constituted 95% (503/532) of the specimens. Overall, 8.1% (43/532) children received potential Dx, and 90% (479/532) received at least one PGx variant. Sequence variants (SV) constituted 70% (30/43) of the Dx (with 1 heteroplasmic mtDNA SV), while 30% (13/43) Dx were CNV analysis assisted calls (4 microdeletions [1q21.2, 16p13.11, 20q13.33 and Xp22.33], a partial *ASH1L* deletion , 4 microduplications [7q11.23, 16p11.2, 17p12, 22q11.2], mosaic 12p triplication, likely der(4)t(4;9)(p16.1;p24.3), mosaic trisomy 8 and uniparental disomy 16). Pathogenic SVs in medically actionable (ACMG) genes constituted 14% (6/43) of Dx (2x*LDLR*, 2x*BTD*, *MYBPC3*, *HNF1A*). Reduced penetrance (RP) variants comprised 30% (13/43) of Dx, including SVs in *TNFRSF13B*, *IFNGR1*, *G6PD*, *SGCE*, *COL4A3*, *PROKR2*, *HNF1A,* and recurrent CNV syndromes (1q21.1 and 16p13.11 deletion, and 16p11.2 duplication syndromes). In one individual 2 RP variants were found (*G6PD* and *COL4A3)*. Other Dx SVs involved genes associated with a range of conditions, including collagenopathies, cancers, eye, heart, or blood defects, as well as skeletal and neurodevelopmental disorders. Our results indicate that WGS screening can serve as a “one-stop shop” for uncovering a wide range of looming genetic conditions in apparently healthy children, this way enabling the family and treating clinicians to take timely appropriate actions to maximize their healthcare outcomes and inform future reproductive decisions.

S72. Sequencing vs panel testing: Is more better?

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 475. An integrated multiomic panel as an excellent tool for the genetic diagnosis of metabolic diseases: Our first 3,720 patients

Authors:

J. Pinto Basto[1], L. S. Almeida[1], C. Pereira[1], R. Aanicai[1], S. Schröder[1], T. Bochinski[1], A. Kaune[1], A. Urzi[1], T. C. L. Spohr[1], N. Viceconte[1], S. Oppermann[1], M. Alasel[1], S. Ebadat[1], S. Iftikhar[1], E. Jasinge[2], S. M. Elsayed[3], H. Tomoum[4], I. Marzouk[5], A. B. Jalan[6], A. Cerkauskaite[7], R. Cerkauskiene[7], T. Tkemaladze[8], A. M. Nadeem[9], I. G. Mahmoud[10], F. A. Mossad[10], M. Kamel[10], L. A. Selim[10], H. A. Cheema[9], O. Paknia[1], C. Cozma[1], C. Juaristi-Manrique[1], P. Guatibonza-Moreno[1], T. Böttcher[1], F. Vogel[1], A. Bertoli-Avella[1], P. Bauer[1]; [1]CENTOGENE, Rostock, Germany, [2]Lady Ridgeway Hosp. for Children, Colombo, Sri Lanka, [3]Med. Genetics Dept., Faculty of Med., Ain Shams Univ., Cairo, Egypt, [4]Dept. of Pediatrics, Faculty of Med., Ain Shams Univ., Cairo, Egypt, [5]Alexandria Univ. Children Hosp., Alexandria, Egypt, [6]Navi Mumbai Inst. of Res. In Mental And Neurological Handicap (NIRMAN) / Pediatric Geneticist, Navi Mumbai, India, [7]Faculty of Med., Vilnius Univ., Vilnius, Lithuania, [8]Dept. of Molecular and Med. Genetics, Tbilisi State Med. Univ., Tbisili, Georgia, [9]Pediatric Gastroenterology, Hepatology and Nutrition, the Children's Hosp. and Inst. of Child Hlth., Lahore, Pakistan, [10]Cairo Univ. Children Hosp. (Abu El Reesh Children's Hosp.), Metabolic, Neurology, Cairo, Egypt

Abstract:

Background/Objectives
To optimize the diagnosis of inherited metabolic disorders (IMDs), we developed a one-test solution. For this purpose, we have implemented a next generation sequencing (NGS) panel with 206 genes that integrates genetic and biochemical testing for an efficient diagnosis of more than 180 metabolic diseases. To our knowledge, there is no other integrated approach currently being offered in clinical practice.
We present our experience of 3,720 patients in our clinical diagnostic setup and validate the use of this multiomic approach as a first-line diagnostic tool for patients suspected of having IMDs.
Methods
The study cohort comprised all consecutive index patients (n=3,720) for whom the panel was performed and reported within a period of 2 years from 62 countries (July 2019-2021). The panel design included 206 genes with single nucleotide and copy number variant (SNV/CNV) detection, followed by semi-automatic variant filtering and reflex biochemical testing (25 assays).
Results
The study represents a global cohort. Over 60% of the patients originated from Africa and Asia (33% and 30%, respectively). Other patients originated from Europe (24%), Latin America (6%), the Middle East (6%), and North America (1%). In 1,389 patients (37%), a genetic diagnosis was achieved. Within this cohort, the highest diagnostic yield was obtained for patients from Asia (57.5%, mainly from Pakistan). Overall, 701 pathogenic/likely pathogenic unique SNVs and 40 CNVs were identified. In 620 patients, the result of the biochemical tests guided variant classification and reporting. Top five diagnosed diseases were: Gaucher disease, Niemann-Pick disease type A/B, phenylketonuria, mucopolysaccharidosis type I, and Wilson disease.
Conclusion
Integrated genetic and biochemical testing facilitated the decision on clinical relevance of the identified variants and led to a high diagnostic yield (37%), which is comparable to exome/genome sequencing. More importantly, up to 43% of the patients (n=610) could benefit from medical treatments (e.g., enzyme replacement therapy). The established genetic diagnosis in nearly 1,400 patients is expected to lead to benefits in treatment and/or counseling for the families. This multiomic approach constitutes a unique and highly effective tool for the genetic diagnosis of IMDs.

S72. Sequencing vs panel testing: Is more better?

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 476. Clinical application of next-generation sequencing for the diagnosis of Mendelian disorders in a highly consanguineous population: results from more than 1400 Iranian families

Authors:

A. Abolhassani[1], Z. Fattahi[1,2], M. Beheshtian[1,2], M. Fadaee[1], R. Vazehan[1], F. Ahangari[1], S. Dehdahsi[1], M. Faraji Zonooz[1], E. Parsimehr[1], Z. Kalhor[1], F. Peymani[1], M. Babanejad[1,2], M. Mozaffarpour Nouri[1], F. Fatehi[3], P. Jamali[4], M. Ashrafi[5], P. Karimzadeh[6], H. Habibi[7], K. Kahrizi[2], S. Nafissi[3], A. Kariminejad[1], H. Najmabadi[1,2]; [1]Kariminejad - Najmabadi Pathology & Genetics Ctr., Tehran, Iran, Islamic Republic of, [2]Genetics Res. Ctr., Univ. of Social Welfare & Rehabilitation Sci., Tehran, Iran, Islamic Republic of, [3]Dept. of Neurology, Neuromuscular Res. Ctr., Shariati Hosp., Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [4]Genetic Counseling Ctr., Shahroud Welfare Organization, Semnan, Iran, Islamic Republic of, [5]Children's Med. Ctr., Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [6]Pediatric Neurology Dept., Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [7]Hamedan Univ. of Med. Sci., Hamedan, Iran, Islamic Republic of

Abstract:

Next Generation Sequencing (NGS) has been proven to be one of the most powerful diagnostic tools for rare Mendelian disorders, especially genetically heterogeneous conditions. While the current diagnostic rate of NGS-based testing in unselected cohorts of patients generally ranges from 25% to 34%, several studies on Middle Eastern populations have reported a higher yield of up to 48%, correlated with a high level of consanguinity and the predominance of autosomal recessive diagnoses. Herein, we evaluated the diagnostic utility of NGS testing across different clinical indications in the previously unstudied consanguineous population of Iran. We report the results of clinical NGS testing in 1436 cases with a wide range of suspected Mendelian diseases. A total of 1075 exome sequencing and 361 targeted gene panel sequencing were performed over an 8-year period at a single clinical genetics laboratory, with the majority of cases tested as proband-only (91.5%). We identified 659 pathogenic or likely pathogenic variants, including 241 novel variants, associated with over 340 known genetic conditions. The overall diagnostic rate was 46.6%, with the highest yields in patients with abnormalities of the skin (67.6%), of blood and blood-forming tissues (64.7%), of the musculature (54.5%), of the skeletal system (51.4%), of the auditory system (50%), of metabolism (48.8%), and of central motor function (46.1%). Dual molecular diagnoses were observed in 2.3% of positive WES cases. The highly consanguineous nature of this cohort led to the diagnosis of autosomal recessive disorders in the majority of patients (79.5%), and allowed us to determine shared carrier status in couples with suspected recessive phenotypes in their deceased child(ren), when no sample was left for direct testing. We also highlight the observations of recessive inheritance of genes previously associated only with dominant disorders, including *DCTN1, KCNC3, BICC1, GLI3, MITF, HARS1,* and *UFSP2*. Finally, we report 94 recurrent recessive pathogenic variants with possible founder effect in Iranian population. This is the most comprehensive view of the mutational spectrum of known Mendelian disease throughout the Iranian population, which can serve as a unique resource for clinical genomic studies locally and beyond. Our cohorts genotypic and phenotypic data can contribute to improved diagnostic yield and variant interpretation within the global community of molecular diagnostic labs.

S72. Sequencing vs panel testing: Is more better?

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 477. Microarray analysis using a comprehensive array with enhanced probe coverage for disease genes and SNP probes

Authors:

W. Bi[1,2], S. Anderson[2], R. Song[2], M. Cooper[2], C. Shaw[1], P. Ward[1,2], A. Breman[1,2], B. Yuan[1,2], S. Cheung[1], L. Vossaert[1,2], Y. Wang[1,2], N. Owen[1,2], J. Smith[1,2], C. Bacino[1,2,3], K. Schulze[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX, [3]Texas Children Hosp., Houston, TX

Abstract:

Background: Disease-causing copy number variants (CNVs) may involve single genes or large genomic segments encompassing multiple genes. While large CNVs spanning over a few hundred kilobases often cause genomic disorders, small CNVs involving a single gene may cause a dominant disorder if the gene is dosage sensitive or they may contribute to a recessive disorder by generating a null allele. Enhancing probe coverage in disease-associated genes/regions in microarrays increases CNV detection sensitivity in these regions. **Method:** Since 2016 chromosomal microarray analysis (CMA) has been provided at Baylor Genetics using an Agilent custom-designed high-resolution oligonucleotide array. This array has exon-targeted probe coverage for >4,200 disease genes associated with dominant or recessive disorders as well as candidate disease genes. In addition, the array contains 60K SNP probes enabling detection of copy number neutral regions of absence of heterozygosity (AOH) and probes for the mitochondrial genome allowing detection of >2 Kb deletions. **Results:** Disease-causing CNVs affecting only single genes were detected in 145 individuals including 134 CNV losses and 11 CNV gains. Among the 91 unique genes involved in single gene CNVs, the most frequently affected genes
were *DMD* (N=17), *TM4SF20* (N=8), *NRXN1* (N=7), *SHANK3* (N=5), *KATNAL2*, *BRCA1* and *USP7* (N=4). The smallest loss is a 0.2 Kb single exon heterozygous deletion in the *TM4SF20* gene, while the smallest gain is a heterozygous 0.3 Kb single exon duplication in the *BRCA1* gene. Biallelic CNVs in autosomal recessive (AR) genes were detected in 10 individuals including 9 homozygous deletions, and one intragenic deletion in *NAXD* in compound heterozygosity with another multiple gene deletion. In addition, one intragenic deletion in *LRBA* was confirmed in compound heterozygosity with a point mutation that was detected by exome. Over 1000 additional deletions involving single AR genes were detected and further studies are needed to know whether they contributed to phenotypes. Uniparental isodisomy for imprinted chromosomes was detected in 9 cases, and AOH >20 Mb in a single imprinted chromosome was detected in 13 cases, one of which was confirmed to be UPD(15). In addition, a ~7 Kb mitochondrial deletion was detected in one case which was confirmed by mitochondrial whole genome sequencing. **Conclusion:** Our data indicated that exon-targeted microarray with enhanced coverage for disease genes and SNP probes is capable of detecting small intragenic copy number changes. CMA using exon-targeted microarray with SNP probes may lead to diagnosis of AR disorders and UPD in addition to making diagnosis of genomic disorders.

S72. Sequencing vs panel testing: Is more better?

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 478. Whole Genome Sequencing Analyses of 45,090 Individuals Reveal Rare Coding and Noncoding Variants Associated with Kidney Function

Authors:

Z. Li[1], X. Li[1], B. Lin[2], H. Kramer[3], N. Franceschini[2], X. Lin[1], TOPMed Kidney Function Working Group; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Div. of Nephrology and Hypertension, Loyola Univ., Chicago, IL

Abstract:

Introduction
Chronic kidney disease (CKD) affects over 850 million adults worldwide and is projected to be the 5th most common cause of years of life lost by 2040. CKD is defined by low estimated glomerular filtration rate (eGFR) and/or increased urine albumin to creatinine ratio (UACR). Genome wide association studies (GWAS) have reported associations of eGFR and UACR with thousands of common and low-frequency variants, but these variants account for only a small fraction of heritability. Rare variants (RVs) may account for some of the unaccounted heritability. Large-scale whole-genome sequencing (WGS) studies, such as the multi-ethnic NHLBI Trans-Omics Precision Medicine (TOPMed) Program, provide the opportunity to assess associations of eGFR and UACR with rare variants across the genome, especially in the noncoding region.
Hypothesis
Rare variant aggregations are associated with eGFR and UACR.
Methods
We applied our newly developed STAARpipeline to detect rare variants (MAF < 0.01) associated with eGFR and UACR using 45,090 and 18,869 individuals from TOPMed Freeze 8 WGS data. STAARpipeline provides gene-centric analysis and non-gene-centric analysis using a variety of coding and noncoding masks. The gene-centric analysis provides five coding and eight noncoding functional categories. The non-gene-centric analysis includes sliding window analysis with fixed sizes and dynamic window analysis with data-adaptive sizes.
Results
For eGFR, the gene-centric analysis identified a genome-wide significant association of missense RVs in *SLC47A1* at the Bonferroni-corrected level 5.00E-07 (=0.05/20,000/5). After conditioning on known eGFR-associated variants, the strength of the association was attenuated but it remained significant at level 2.50E-06. For UACR, the 2-kb sliding window procedure identified a genome-wide significant association of RVs in an intergenic region near *ASB1* at the Bonferroni-corrected level 1.88E-08 (=0.05/2.66E06). After conditioning on known UACR-associated variants, the association remained significant at the same level 1.88E-08. The dynamic window procedure additionally detected two significant associations at the genome-wide error rate 0.05 level, including intronic RVs of *TRIM67* and *ERCC6L2*. These two associations remained significant at level 1.88E-08 in conditional analysis.
Conclusions
Four new RV associations, including missense RVs in *SLC47A1* with eGFR, RVs in an intergenic region near *ASB1* with UACR, and intronic RVs of *TRIM67* and *ERCC6L2* with UACR, were identified using the TOPMed WGS Freeze 8 data through STAARpipeline. These findings suggest a role of rare variants in kidney traits.

S72. Sequencing vs panel testing: Is more better?

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 479. Diagnostic yield of pediatric and prenatal exome sequencing in a diverse population

Authors:

A. Slavotinek[1], S. Rego[2], N. Sahin-Hodoglugil[3], M. Kvale[3], T. Yip[4], H. Hoban[4], B. Lianoglou[4], S. Outram[4], B. Anguiano[4], F. Chen[4], C. Curry[5], R. Gallagher[4], M. Gardner[6], B. Mendelsohn[7], E. Sherr[4], J. Shieh[4], J. Strober[4], A. Tam[4], J. Tenney[4], A. Whittle[8], Y. Mavura[4], J. Van Ziffle[4], P. Devine[4], U. Hodoglugil[4], P-M. Martin[4], T. Sparks[4], B. Koenig[4], S. Ackerman[4], N. Risch[4], P-y. Kwok[4], M. Norton[4]; [1]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [2]Univ. of California, San Francisco, Dept. Pediatrics, San Francisco, CA, [3]Inst. for Human Genetics, Univ. of California, San Francisco, San Francisco, CA, [4]Univ. of California, San Francisco, San Francisco, CA, [5]Univ California, San Francisco/ Fresno, Fresno, CA, [6]UCSF Benioff Children's Hosp. Oakland, Oakland, CA, [7]Kaiser Permanente, Oakland, CA, [8]Zuckerberg San Francisco Gen. Hosp., San Francisco, San Francisco, CA

Abstract:

We evaluated the diagnostic yield of exome sequencing (ES) in a cohort of predominantly underrepresented minority (URM) and underserved (US) patients suspected to have a genetic disorder. Our unique cohort included both prenatal and pediatric patients that were all analyzed with the same ES pipeline, thus enabling a direct comparison between the two groups. Pediatric patients (n = 529) had neurocognitive disabilities or multiple congenital anomalies and prenatal patients (n = 316) had one or more structural anomalies, disorders of fetal growth, or fetal effusions. 554/845 (65.6%) of the total cohort had at least one parent who self-identified as URM, 155/845 (18.3%) were non-URM (i.e. white/European), and 136/845 (16.1%) had unknown race/ethnicity. 457/529 (86.4%) of pediatric families and 146/316 (46.2%) of prenatal families met one of the US categories, defined by MediCal health insurance and domicile in a medically underserved area and/or health professional shortage area. We identified pathogenic or likely pathogenic variants in 202/845 (23.9%) of patients, with a higher diagnostic rate in pediatric patients (26.8%) compared to the prenatal group (19.0%; p = .01). The frequency of inconclusive results was also greater in the pediatric patients (13.8%) compared to the prenatal patients (6.3%; p = .0008). Of interest, the higher diagnostic yield in the pediatric cases was largely attributable to pathogenic variants in genes with autosomal dominant inheritance. In pediatric families, there was a similar diagnostic yield for URM and non-URM individuals (26.4% vs 30.3%; p = .51). The diagnostic rate was also not significantly different for US families compared to non-US families (26.0% vs 31.9%; p = .091). Inconclusive results were present in 17.1% of URM patients, compared to 7.6% of non-URM patients (p = .123), and in 14.7% of US patients and 8.3% of non-US patients (p = .148). For the prenatal patients, the diagnostic yield was also similar between URM (17.1%) and non-URM (15.9%) families and for US (17.8%) versus non-US (20.0%) families. Likewise, the rates of inconclusive results were similar, with 7.6% for URM versus 3.4% for non-URM families and 7.5% for US versus 5.3% for non-US families of prenatal patients. Therefore, in both pediatric and prenatal patients, the diagnostic yield and the rate of inconclusive results were not significantly different in the offspring of parents who self-reported URM status compared to those who self-reported non-URM status or for those with US vs non-US status. Importantly, our data support the application of genomic sequencing in patients from diverse population groups with referral indications for ES.

S73. Structural variation in population and disease

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 482. Resolving the exact breakpoints and sequence rearrangements of large neuropsychiatric copy number variations (CNVs) at single base-pair resolution using CRISPR-targeted ultra-long read sequencing (CTLR-Seq)

Authors:

B. Zhou[1], G. Shin[1], L. Vervoort[2], S. Greer[1], Y. Huang[1], R. Pattni[1], J. Vermeesch[2], H. Ji[1], A. Urban[1]; [1]Stanford Univ. Sch. of Med., Palo Alto, CA, [2]KU Leuven, Leuven, Belgium

Abstract:

Large copy number variants (CNVs) such as those in 16p11, 22q11, 1q21, and 15q13, i.e. recurrent heterozygous chromosomal aberrations in the form of large deletions and duplications, are the highest known risk factors for psychiatric disorders. Highly complex and repetitive human-specific sequences >100kb to megabases i.e. segmental duplications (SegDups) act as substrates that trigger such rearrangements. These SegDups contain functional sequence elements and have so far been impenetrable to standard DNA sequencing analysis. Thus, the exact genomic boundaries and breakpoint locations of large neuropsychiatric CNVs remain unknown.

We developed a novel approach (CRISPR-targeted-ultra-long-read sequencing, CTLR-Seq) where we combine Cas9-targeting, pulse-field gel electrophoresis, and ultra-long nanopore sequencing as well as linked-read sequencing, to isolate intact large CNV rearrangements in a haplotype-specific fashion to resolve their content and rearrangement sequences using de novo assembly for the major neuropsychiatric CNVs. CTLR-Seq was also applied to the parent of origin of 22q11.2 patients to map the precise breakpoints of the typical 22q11.2 deletion. Fiber-FISH was used to guide and confirm the 22q11.2 rearrangements.

We applied CTLR-Seq and, for the first time, sequence resolved the major neuropsychiatric CNVs, 16p11.2 deletion (n=4), 16p11.2 duplication (n=8), 15q13.3 deletion (n=4), 15q13.3 duplication (n=3), 1q21 deletion (n=4), 22q11.2 deletion (n=4), mapping out their exact sequence content and rearrangement structure. In 3 of the 4 individual 22q11 patients harboring the typical 3Mbp deletion, we also applied CTLR-Seq to their respective parent of origin to completely resolve the SegDup sequences prior to rearrangement and also for the first time, mapping the typical 22q11 deletion breakpoint in these individuals. These three typical 22q11 deletion breakpoint locations correspond to different retrotransposon (LINE, Alu, and HERV) elements, all within GGT2, suggesting that the mechanism of formation for the typical 22q11.2 deletion is via transposon-mediated recombination and that GGT2 is a recombination hotspot.

Our development of CTLR-Seq allows for the complete and haplotype-specific resolution of large neuropsychiatric CNVs rearrangements and their breakpoints at single base-pair resolution. This will make it possible to analyze cohorts of CNV patients to discover SegDup rearrangement variations in different patients, breakpoint locations, and mechanisms of formation, to investigate a possible role of such variance as genetic modifier, to determine predisposing haplotypes to CNV formation.

S73. Structural variation in population and disease

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 483. Structural variation analysis on 16,905 whole-genome sequencing data from Alzheimer's Disease Sequencing Project (ADSP)

Authors:

H. Wang[1,2], B. Dombroski[1,2], B. Vardarajan[3,4], A. Tucci[5], J. Farrell[6], C. Dalgard[7], L. Farrer[6,8,9,10,11], J-Y. Tzeng[5], Y. Leung[1,2], L-S. Wang[1,2], G. Schellenberg[1,2], W-P. Lee[1,2]; [1]Dept. of Pathology and Lab. Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [2]Penn Neurodegeneration Genomics Ctr., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [3]Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, Coll. of Physicians and Surgeons, Columbia Univ., New York, NY, [4]Dept. of Neurology, Coll. of Physicians and Surgeons, Columbia Univ. and the New York Presbyterian Hosp., New York, NY, [5]Bioinformatics Res. Ctr., North Carolina State Univ., Chapel Hill, NC, [6]Dept. of Med. (BioMed. Genetics), Boston Univ. Sch. of Med., Boston, MA, [7]Dept. of Anatomy, Physiology & Genetics, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [8]Dept. of Neurology, Boston Univ. Sch. of Med., Boston, MA, [9]Dept. of Ophthalmology, Boston Univ. Sch. of Med., Boston, MA, [10]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [11]Dept. of Epidemiology, Boston Univ. Sch. of Publ. Hlth., Boston, MA

Abstract:

Background: Genetic variation within common genome-wide associated loci does not account for a substantial proportion of heritability of Alzheimer's disease (AD). Structural variation (SV), including copy number variations (CNVs), contributes to the genetic basis of human disease and is likely a potential source of "missing heritability" in AD. Leveraging whole-genome sequencing (WGS) data for 16,905 samples collected by 17 studies from The Alzheimer's Disease Sequencing Project (ADSP) R3 dataset, we systematically investigated the contribution of SVs in the genetic architecture of AD.
Methods: Two commonly used SV calling algorithms (Manta and Smoove) were applied to generate an individual call set for each sample. Graphtyper2 was then used for joint-genotyping to improve sensitivity, reduce false positives, adjust SV breakpoints, and merge call sets. SVs were functionally annotated by VEP and AnnoSV. To investigate risk SVs potentially associated with AD, we identified SVs in linkage disequilibrium (LD) with AD SNVs from GWAS Catalog and SVs on 20 AD pathogenic genes. In addition, we performed an association analysis of common and rare SVs with AD status.
Results: After joint-genotyping, we detected a total of 400,234 SVs (231,385 deletions, 45,839 duplication, 119,648 insertions and 3,362 inversions) from 16,905 samples. Compared to common variants, rare variants (especially singletons) tend to be more deleterious as determined by a higher portion in regulatory/coding regions versus non-functional regions. Overall, we detected a higher burden of common CNVs (OR = 1.10, $P < 0.001$) and rare CNVs (OR = 1.15, $P < 0.001$) in cases compared to controls. A rare deletion (AF = 3.11×10^{-4}, chr2:105731359-105736864) is in LD (R^2=0.44) with recent identified rare risk loci (rs115186657/rs143080277) in *NCK2*. We identified 100 SVs on the AD pathogenic genes, including one deletion in *APOE* exon 4 and 12 rare variants that may cause frameshift or transcript amplification/ablation. Finally, 73 common and 4 rare SVs with an FDR < 0.05 were identified by association analysis, particularly SVs in *NAALADL2*, *CFAP99* and *PDE4D* are candidates for further experimental validation.
Conclusion: We composed an integrative pipeline for SV calling and provided a comprehensive catalogue of SVs detected on ADSP R3 17K WGS data. AD associated rare SVs with high functional impact were identified, indicating a significant role for SVs in the genetic basis of AD.

S73. Structural variation in population and disease

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 484. Detection of complex structural genome variants using ARC-SV and their enrichment inside genes of neurodevelopmental pathways

Authors:

C. Hughes[1], B. Zhou[1], J. Arthur[2], T. Kim[3], G. Song[4], D. Palejev[5], W. Wong[2], A. Urban[1]; [1]Dept. of Psychiatry and Behavioral Sci., Dept. of Genetics, Stanford Univ. Sch. of Med., Stanford, CA, [2]Dept. of Statistics, Stanford Univ., Stanford, CA, [3]Sch. of Computer Sci. and Engineering, Coll. of Engineering, Pusan Natl. Univ., Busan, Korea, Republic of, [4]Sch. of Computer Sci. and Engineering, Coll. of Engineering, Pusan Natl. Universit, Busan, Korea, Republic of, [5]GATE Inst., Sofia Univ., Inst. of Mathematics and Informatics, Bulgarian Academy of Sci., Sofia, Bulgaria

Abstract:

Complex structural variants (cxSVs), or genomic alterations affecting ≥ 50 bp that are not reducible to a single rearrangement event (e.g. duplication, deletion, inversion, etc.), are major components of variation in human genomes. Due to the technological challenges in identifying and resolving their complex structures, cxSVs have been largely excluded from population-scale association studies. Large genomic alterations, such as those constituted by cxSVs, have previously been linked to psychiatric phenotypes like schizophrenia and ASD.

To overcome the technical hurdles with detecting cxSVs from standard whole-genome sequencing, we developed and applied a novel probabilistic algorithm “Automated Reconstruction of Complex Variants” (ARC-SV) to detect and catalog cxSVs in 4,293 high-coverage whole genomes representing 213 diverse human populations. We then performed functional annotation and pathway enrichment analysis by hypergeometric testing and using FDR correction. To assess if cxSV enrichment was spurious, we conducted pairwise testing of gene length, probability of being loss-of-function intolerant (pLI), variant counts, and semantic similarity between the whole genome, cxSVs, and simple structural variants (SVs).

We identified 8,130 unique cxSVs across the 4,293 human genomes. On average, each genome contained 129 cxSVs with a typical size of 7.1 kb. Of those 8,130 variants, 3,172 reside in 1,876 unique genes for which we find neurodevelopmental pathway enrichment alongside overrepresentation in nervous-system tissues. cxSVs are at 1.28 greater odds of gene overlap compared to simple SVs. In addition, cxSVs are associated with differentially upregulated genes during brain development at ages 21 weeks gestation, 1 year, and 23 years. Significant cxSV ontology terms include: synapse, cell projection, generation of neurons, ASD, substance abuse, schizophrenia, depression, and cognitive ability. Associations attenuate if singleton cxSVs are excluded suggesting that individually private cxSVs are an important contributor to neural phenotypes.

cxSVs and SVs share 0.79 semantic similarity across molecular, cellular, and biological ontologies. However, the semantic similarity of disease ontologies is 0.15 with cxSVs being exclusively neuropsychiatric.

We have identified thousands of novel genomic cxSVs from diverse human populations and find that these cxSVs are enriched in neural pathways. As most cxSVs analyzed here are private to individuals, de novo cxSVs may preferentially contribute to individual neural phenotypes and therefore be leveraged in precision medicine and nosological subtyping.

S73. Structural variation in population and disease

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 485. Advances in long-read sequencing and telomere-to-telomere assembly enable discovery of cryptic ring chromosome breakpoints and highlight complex rearrangements of acrocentric p-arms

Authors:

Y. Mostovoy[1], P. M. Boone[1], Y. Huang[2], K. V. Garimella[2], F. Quintero-Rivera[3], B. E. Russell[4], T. Bowers[2], T. DeSmet[2], S. Gabriel[2], J. Gusella[1], H. Brand[1], M. E. Talkowski[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst., Cambridge, MA, [3]Univ. of California Irvine (UCI), Irvine, CA, [4]UCLA Sch. of Med., San Luis Obispo, CA

Abstract:

Resolving genomic variation at sequence resolution within highly repetitive and complex regions of the human genome has been largely intractable with conventional genomic tools. As a consequence, there is a dearth of sequence-based information about genomic events such as formation of repeat-associated ring chromosomes, Robertsonian translocations, or population variation across the p-arms of acrocentric chromosomes. Delineation of these events at sequence resolution could advance our understanding of their mechanism of formation and functional impact. We reasoned that two recent technical milestones could be leveraged to gain insight into the sequence of these still largely uncharacterized events: 1) maturation of long-read sequencing technologies (e.g. Pacific Biosciences [PacBio] and Oxford Nanopore [ONT]); and 2) the Telomere-to-Telomere (T2T) gapless genome assembly. We performed ONT sequencing on DNA from 12 patient-derived cell lines for which karyotyping identified likely complex rearrangements including: six ring chromosomes, three Robertsonian translocations, two inter-chromosomal translocations, and one pericentric inversion. Among these rearrangements, 75% involved acrocentric p-arms. Reads were aligned to the T2T assembly and breakpoints determined using copy-number variant detection (GATK-gCNV), breakend capture (pbsv), and custom scripts and visualization. We discovered and resolved rearrangements for 9 of the 12 samples, including all ring chromosomes, one Robertsonian translocation, and 6/9 samples involving acrocentric p-arms. Strikingly, two samples - one ring and the pericentric inversion - had breakpoints inside tandem arrays of ribosomal DNA. Of the rings, two samples had telomeric repeats at the breakpoint, while four involved terminal deletion of both chromosome arms. Complex breakpoint structures were seen in three of the rings, including an inverted translocation and an inter-chromosomal insertion. A Robertsonian translocation had an insertion from chr19 mediating the fusion of 14p and 15p. To further assess the structure of acrocentric p-arms, we analyzed PacBio HiFi data of 50 normal samples and observed multiple large-scale (>1Mb) rearrangements including translocations and inversions. In summary, integrating long-reads with the T2T assembly enabled detection and precise characterization of chromosomal rearrangements in previously intractable regions and illuminates a wide range of structural diversity and mechanisms. The approaches presented here can be used to probe a variety of complex structural variants with important implications for molecular diagnosis and genome biology.

S73. Structural variation in population and disease

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 486. Pangenome graphs for the analysis of rare genetic diseases

Authors:

C. Groza[1], W. Cheung[2], G. Bourque[1], T. Pastinen[2]; [1]McGill Univ., Montreal, QC, Canada, [2]Children's Mercy Kansas City, Kansas City, MO

Abstract:

Human genomes accessed by long-read genome sequencing using PacBio Sequel IIe enable variant calling by de novo assembly and graph-based analyses. However, interrogation of differences across such graph structures and understanding their potential function requires large collections of high quality assemblies and projections to known annotations. Leveraging 252 haplotype-resolved assemblies from 123 rare disease families, included in the “Genomic Answers for Kids” (GA4K) program at Children’s Mercy Kansas City, we built a large pangenome graph to detect rare structural variation (SV) and gain first insights into the utility of personal assemblies and graph genomes in unsolved rare diseases. We started by generating haplotype-resolved assemblies using hifiasm with graph trio binning. Next, we ran minigraph to construct a pangenome graph by first augmenting the T2T-CHM13v2 backbone reference with 94 assemblies released by the HPRC and then adding 252 haplotype-resolved assemblies from the rare disease families that are at least 2.7 gigabases long, have at most 3000 contigs and a minimum NG50 of 10 megabases. Overall, we identified 144217 bubbles in the graph that were larger than 50 bp, of which 48120 were specific to our cohort. One-third (12623/39426) of rare bubbles (appearing in 12 or fewer haplotypes) were also found to be overlapping a rare structural variant (SV) called by pbsv from HiFi-CCS read alignments against GRCh38. We intersected bubbles with the ENSEMBL annotation of T2T-CHM13v2 and found that ultra-rare (singleton) variants observed in patients overlapped conserved coding sequences (CCDS) 2810 times, including 657 alleles in OMIM disease genes. Among these OMIM CCDS singleton variants, manual curation suggests that 54% are true SVs (most remaining are likely errors in primary assemblies), but up 20% can be missed by reference based SV calling.
These results indicate that on average, one of three singleton SVs, potentially disrupting OMIM gene coding sequence, can be detected only by assembly-based methods. We are developing scoring for de novo primary assembly quality and validation by primary read alignments at rare variant sites. We will use this to robustly extract and annotate alleles that are relevant to rare disease in pangenomes.

S73. Structural variation in population and disease

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 487. Human Y chromosome - *de novo* assembly and comprehensive analysis of genetic variation across 45 diverse haplotypes

Authors:

P. Hallast[1], P. Ebert[2], F. Yilmaz[1], M. Loftus[3], P. A. Audano[1], W. Zhou[4], G. Logsdon[5], W. Höps[6], C. Li[7], M. Bonder[8], P. C. Dishuck[5], D. Porubsky[5], R. Snajder[8], Q. Zhu[1], K. Munson[5], P. Hasenfeld[9], W. T. Harvey[5], Human Genome Structural Variation Consortium (HGSVC), J. O. Korbel[9], E. E. Eichler[5,10], C. Beck[1,11], X. Shi[7], M. K. Konkel[3], T. Marschall[2], C. Lee[1]; [1]The Jackson Lab. for Genomic Med., Farmington, CT, [2]Heinrich Heine Univ., Düsseldorf, Germany, [3]Clemson Univ., Clemson, SC, [4]Univ. of Michigan Med. Sch., Ann Arbor, MI, [5]Univ. of Washington Sch. of Med., Seattle, WA, [6]EMBL Heidelberg, Heidelberg, Germany, [7]Temple Univ., Philadelphia, PA, [8]German Cancer Res. Ctr., Heidelberg, Germany, [9]European Molecular Biology Lab. (EMBL), Heidelberg, Germany, [10]Howard Hughes Med. Inst., Seattle, WA, [11]10The Univ. of Connecticut Hlth.Ctr., Farmington, CT

Abstract:

The complex sequence composition of mammalian Y chromosomes, including dense repeats and large ampliconic regions, has led to their systematic omission from genomic analyses. The male-specific Y is one of the smallest chromosomes in the human genome, yet more than 50% of its ~57 Mbp content is missing from the GRCh38 reference sequence. For the first time, a complete human chromosome Y has been assembled by the Telomere-to-Telomere (T2T) Consortium. For a more holistic understanding of Y chromosome evolution, we assembled and analysed 45 Y chromosomes from 1000 Genomes Project males. Using combined high-coverage PacBio HiFi and Oxford Nanopore ultra-long sequence data from the Human Genome Structural Variation Consortium (HGSVC) and Human Pangenome Reference Consortium (HPRC), we implemented the Verkko pipeline to produce complete and highly contiguous Y-chromosomal assemblies. The samples selected for our analysis offer maximal representation of human Y-chromosomal diversity, including major African and non-African Y haplotypes, with the root of the tree dated to approximately 185 thousand years ago (kya) (95% HPD interval: 162-212 kya), thus covering genetic variation across a substantial period of modern human Y chromosome evolution. We describe inter-Y chromosome variation at the base-pair level for the first time, including the large Yq12 heterochromatic region and the centromere, both largely absent from GRCh38. The total Y chromosome size varies up to twofold between individuals, mostly due to differences in the Yq12 heterochromatic block. Additional heterochromatic regions, including the centromere, show local patterns of evolution, including duplications and deletions. These variants lead to bursts of Y-chromosomal expansions and contractions over short periods of time and contribute to substantial differences between samples. We report extensive structural variation across the studied males, with >60% of Y-chromosomal euchromatic content being affected by large and often recurrent inversions up to multi-Mbp in size, with only the most closely related Y chromosomes showing similar composition. Our study is the first comprehensive analysis of the full extent of Y chromosome variation across multiple males, establishing a detailed blueprint of Y-chromosomal sequence identity that has the potential to form the core dataset for multiple and multidisciplinary follow-up studies.

S74. Challenges in everyday clinical genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 490. Costs and diagnostic yield of whole genome sequencing in neurodevelopmental disorders

Authors:

M. Johansson Soller[1], H. Runheim[2], A. Hammarström[1], M. Pettersson[1], A. Lindstrand[1], L-Å. Levin[2]; [1]Karolinska Hosp., Stockholm, Sweden, [2]Linköping Univ., Linköping, Sweden

Abstract:

Background: Whole genome sequencing (WGS) has the potential to be a comprehensive genetic test, especially relevant for individuals with neurodevelopmental disorders, syndromes and congenital malformations. However, the cost consequences of using WGS as a first-line genetic test for these individuals are not well understood. The study objective was to compare the healthcare costs and diagnostic yield when WGS is performed as the first-line test instead of chromosomal microarray analysis (CMA).Materials and Methods: Two cohorts were analyzed retrospectively using register data, cohort CMA (418 patients referred for CMA at the department of Clinical Genetics, Karolinska University Hospital, during 2015) and cohort WGS (89 patients included in a WGS-first prospective study in 2017). The analysis compared healthcare consumption over a two-year period after referral for genetic testing and diagnostic yield over a two- and three-year period after referral.Resluts: The mean healthcare cost per patient in cohort WGS was $2,339 lower compared to cohort CMA ($-2,339, 95%CI, -12,238-7,561; P=0.64) including higher costs for genetic investigations ($1,065, 95%CI, 834-1,295; P<0.001) and lower costs for outpatient care ($-2,330, 95%CI, -3,992-(-669); P=0.006). The diagnostic yield was 23% higher for cohort WGS (cohort CMA 20.1%, cohort WGS 24.7%) (0.046, 95%CI, -0.053-0.145; P=0.36).Conclusions: WGS as a first-line diagnostic test for individuals with neurodevelopmental disorders is associated with statistically non-significant lower costs and higher diagnostic yield compared with CMA. This indicates that prioritizing WGS over CMA in health care decision making will yield positive expected outcomes as well as showing a need for further research.

S74. Challenges in everyday clinical genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 491. Insurance denials and diagnostic rates in a pediatric genomic research cohort

Authors:

T. Zion, C. D. Berrios, A. S. A. Cohen, L. Bartik, E. T. Rush, R. Moore, E. Boillat, R. L. Biswell, D. A. Louiselle, L. M. B. Puckett, S. Beyer, S. H. Neal, V. Sierant, M. McBeth, B. Belden, A. M. Walter, M. Gibson, W. A. Cheung, J. J. Johnston, I. Thiffault, E. G. Farrow, E. Grundberg, T. Pastinen; Children's Mercy Hosp., Kansas City, MO

Abstract:

The clinical utility and diagnostic yield of exome sequencing (ES) and multi-gene panel tests in patients with suspected genetic conditions is well-established. Despite this, many patients continue to face barriers to clinical genetic testing through insurance denials. In most cases, it is impossible to know how many patients whose clinical genetic testing is denied by insurance may have an actionable variant because they never receive access to testing. As a result, little research has approached this issue. At the Children’s Mercy Research Institute, we have a unique opportunity to study a subset of patients enrolled in the Genomic Answers for Kids (GA4K) research study, providing low barrier and no-cost access to genetic testing for participants. To date, there are 3,406 clinically symptomatic individuals enrolled in GA4K. We analyzed ES and, in negative cases, genome sequencing (GS) data in the first 1,332 enrolled patients and reviewed their clinical testing history. We identified participants with no previous genetic testing (N=182), testing ordered in parallel to nomination to the study (N=225), previous negative genetic testing (N =679), previous positive genetic testing (N=239), and cases in which past testing history was unclear (N=7). We focused on patients with no previous clinical genetic testing, or previous negative genetic testing at the time of study entry for the purposes of our investigation (N=869) and utilized retrospective reviews of patient charts to determine if insurance denials occurred. Cases were cross-referenced to study results to determine rate of diagnostic findings and findings of interest not meeting diagnostic criteria. A total of 140 cases (16.11%) had insurance prior-authorization denials on file. Of these 140 cases, 13 had clinically diagnostic genetic findings identified through our study, 2 had a partial diagnosis (a second diagnosis is suspected and we identified 1 diagnosis), 20 had strong candidate variants in genes of uncertain significance, 27 had strong candidate variants of uncertain significance in clinically established disease-causing genes, 13 had a partial genotype identified (1 variant in an autosomal recessive condition), and 65 were negative for any strong candidate findings. In total, 75 cases (53.6%) had a pertinent genetic finding. Of the 15 cases with at least 1 diagnostic finding, 6 cases (40%) had clinically actionable results that changed medical management.

S74. Challenges in everyday clinical genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 492. Whole genome germline sequencing: Its role in general health screening in family practice, the first study in the UK

Authors:

A-B. Jones[1], G. Pichert[2], L. Side[3], T. Homfray[4], M. Murray[5], S. Hicks[5], E. Bancroft[6], Z. Kote-Jarai[7], M. Sandberg[5], R. Eeles[7]; [1]The Inst. of Cancer Res./Royal Marsden Hosp., London, United Kingdom, [2]Zurich and 90 Sloane Street, London, United Kingdom, [3]Univ. of Southampton and 90 Sloane Street, London, United Kingdom, [4]Gen. and Cardiac and Prenatal Geneticist; St George's Hosp., Kings Coll. Hosp. and 90 Sloane Street, London, United Kingdom, [5]90 Sloane Street, London, United Kingdom, [6]Royal Marsden Hosp, Sutton, United Kingdom, [7]The Inst. of Cancer Res., London, United Kingdom

Abstract:

Background: Whole genome sequencing (WGS) is now possible due to improvements in next generation sequencing technology. This gives the opportunity to incorporate genetic data into family practice health screening. The fulfilment of the potential of WGS lies in its future implementation as an additional preventative risk identifier either independently or pathway to undertake WGS for actionable findings and assess its potential role in enhancing health screening. Methods: A pilot study comprised of 100 healthy participants from 2020-2022, recruited from a family practice, consented to routine baseline blood tests, cardiac assessment (ECG and echocardiogram), abdominopelvic ultrasound, all performed at the practice. Review of past medical, and family history was done from electronic health records. Germline genetic testing consisted of 84 cancer and 77 cardiac genes and whole genome sequencing (566 actionable genes reported), including higher penetrance monogenic mutations, recessive carrier alterations and pharmacogenomics. A multidisciplinary clinical team reviewed integrated results through an iterative process and fed back to participants. Results: 28 of 100 individuals (28%) had an actionable genetic variant in either cancer (10 individuals), thrombophilia (6), or cardiac rhythm disorder genes (2). Over fifty percent of the participants had a variant in an autosomal recessive gene. Pharmacogenomics results yielded significant variants in over a third of participants. Polygenic Risk Score (PRS) analyses for selected common cancers are in progress. Conclusion: Nearly one third of patients had a significant change in health management for themselves and over 80% for their families. Whole genome sequencing as part of health screening in family practice is feasible and is likely to have significant beneficial health management implications. This study is the first of its kinds in a UK family practice and shows how this can be implemented.

S74. Challenges in everyday clinical genetics

Location: Conv Ctr/Petree C/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 493. The ClinGen General Inborn Errors of Metabolism Gene Curation Expert Panel: Assessing the clinical validity of genes implicated in metabolic disorders

Authors:

W. Craigen[1], E. Brown[2], W. Hankey[3], S. Kathirgamanathan[4], J. Lowry[5], J. Ross[3], R. Fernandez[6], M. Weaver[6], H. Baudet[3], A. Feigenbaum[7], C. Hung[8], E. Spector[9], **J. Goldstein**[3]; [1]Baylor Coll. of Med., Houston, TX, [2]Nutriplexity, Bremerton, WA, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Univ. of Toronto, Toronto, ON, Canada, [5]Univ. of Utah, Salt Lake City, UT, [6]American Coll. of Med. Genetics and Genomics, Bethesda, MD, [7]Univ. of California San Diego, La Jolla, CA, [8]Invitae, San Francisco, CA, [9]Children's Hosp. Colorado, Aurora, CO

Abstract:

The Clinical Genome Resource (ClinGen; www.clinicalgenome.org) is an NIH-funded initiative dedicated to creating a public resource that defines the clinical relevance of genes and variants. ClinGen's Inborn Errors of Metabolism Clinical Domain Working Group (IEM CDWG) is tasked with assembling expert panels, groups of individuals with expertise in specific genes and diseases, to classify gene-disease clinical validity and variant pathogenicity for metabolic disorders. Gene-disease clinical validity classification is carried out by Gene Curation Expert Panels (GCPEs) using the ClinGen clinical validity framework. Clinical validity classifications and the data supporting those classifications are published on the ClinGen website. GCEPs formed under the umbrella of the ClinGen IEM WG include the Fatty Acid Oxidation Disorders, Aminoacidopathies, Peroxisomal Disorders, Mitochondrial Diseases, Monogenic Diabetes, and Lysosomal Diseases GCEPs, with plans to initiate a Congenital Disorders of Glycosylation GCEP in the near future. Each of these GCEPs focuses on a specific subgroup of disorders associated with a large number of genes (>30). In addition to these specific GCEPs, the IEM CDWG recently formed a General IEM GCEP that has 3 main goals:1) To classify the clinical validity of genes associated with IEMs on the Recommended Uniform Screening Panel for newborns (RUSP) that have not already been classified by other GCEPs.2) To classify the clinical validity of genes involved in subgroups of metabolic disorders associated with small numbers of genes (<10) that would not warrant formation of a specific GCEP.3) To classify the clinical validity of genes that do not currently come under the purview of another GCEP but for which there is a need for the curation to be completed. 4) To re-curate gene-disease clinical validity classifications on a regular basis for IEM GCEPs that have completed curation and are no longer active. The General IEM GCEP is currently composed of 5 curators (2 of whom are volunteers), 5 metabolic experts, and 2 coordinators. To date, we have classified the clinical validity of 17 gene-disease pairs, including genes causing conditions on the RUSP, porphyria, galactosemia, bile acid synthesis, and copper metabolism, with plans to curate genes involved in glycogen storage disorders in the near future. We will present the results of our curation efforts to-date, and discuss the challenges of classifying gene-disease clinical validity for metabolic disorders, in particular those disorders with a biochemical phenotype but not necessarily a clinical phenotype.

S75. Gee, What A Session! (GWAS)

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 496. Multi-ethnic GWAS meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease that define new disease subtypes, mechanisms, and predict advanced liver disease

Authors:

X. Du[1], Y. Chen[1], A. Kuppa[1], V. Chen[1], N. Palmer[2], E. K. Speliotes[1], the GOLD Consortium; [1]Univ. of Michigan, Ann Arbor, MI, [2]Wake Forest Sch. of Med., Winston-Salem, NC

Abstract:

Background: Non-alcoholic fatty liver disease (NAFLD) is a heritable condition that is now the most common cause of chronic liver disease in the United States. Its prevalence varies by ancestry. Some genetic variants have been linked to NAFLD but many remain to be identified. Methods: We conducted the largest cross ancestry genome-wide association meta-analysis of imaging and ICD based NAFLD in five cohorts: (1) Genetics of Obesity-related Liver Disease (GOLD) Consortium, consisting of 9 United States and European cohorts with computed tomography (CT)-measured hepatic steatosis; UK Biobank participants with hepatic steatosis defined by magnetic resonance imaging (2) or diagnosis codes (3); (4) eMERGE; and (5) FinnGen. Our meta-analysis included 66,814 individuals with hepatic steatosis defined by imaging and 3,584 cases (vs. 621,081 controls) of NAFLD defined by diagnosis codes. We identified 17 genetic variants associated with hepatic steatosis at genome-wide significance ($p<5x10^{-8}$). We generated a PRS based on the dosage for each genome-wide significant allele, weighted by effect size and evaluated its impact in NAFLD, cirrhosis, and HCC in the Michigan Genomics Initiative (MGI). Phenome-wide association study (PheWAS) was conducted on the genome-wide significant variants in UKBB, followed by k-means clustering of the results. We carried out 2 sample Mendelian Randomization (MR) analyses to define causal phenotypes. We analyzed ancestry-specific allele frequency and effect size in the GOLD Consortium. Results: We identified 17 alleles associated with NAFLD at genome-wide significance ($p<5x10^{-8}$). A higher NAFLD PRS using these variants associated with NAFLD in MGI: odds ratio (OR) 2.83 (95% confidence interval [CI] 2.39-3.34). Higher NAFLD PRS was also associated with increased odds of cirrhosis: top 10% OR 2.47 , 5% 3.39 , and 1% 4.87 and HCC: top 10% OR 2.91 , 5% 4.35, and 1% 6.34 (all $p < 0.05$). PheWAS identified 6 clusters of genes that identify disease subtypes and suggest different mechanisms of disease. MR analyses showed causal effects of hepatic steatosis on cirrhosis and esophageal varices and BMI and waist circumference on hepatic steatosis. Across ancestries in GOLD we found >10% absolute difference in effect allele frequencies (EAF) for NAFLD-increasing variants in *PNPLA3*, *GCKR*, *TRIB1*, and *ADH1B* but there was not heterogeneity of effect size. Conclusions: We identified genetic variants that define new subtypes of NAFLD, their mechanism, and show that they can identify individuals at risk of advanced liver disease. This advances our understanding of NAFLD and opens up new possibilities for diagnosis and treatment.

S75. Gee, What A Session! (GWAS)

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 497. Genome-wide meta-analysis identifies novel risk variants and enables polygenic prediction of preeclampsia and gestational hypertension

Authors:

M. Honigberg[1,2], B. Truong[2], S. Koyama[2], A. Patel[1,2], S. Cho[2], M. Trinder[3], S. Urbut[1,2], S. Patil[4], S. Zollner[4], T. Vy[5], G. Nadkarni[5], R. Do[5], Estonian Biobank Research Team, T. Laisk[6], Genes & Health Research Team, D. van heel[7], P. Natarajan[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Univ. of British Columbia, Vancouver, BC, Canada, [4]Univ. of Michigan, Ann Arbor, MI, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Inst. of Genomics, Univ. of Tartu, Estonia, Estonia, [7]Barts and The London, London, United Kingdom

Abstract:

Preeclampsia and gestational hypertension together affect up to 15% of child-bearing women and are associated with adverse maternal and offspring outcomes. The pathophysiology of these conditions remains incompletely understood, and current tools for prediction, prevention, and treatment are limited. Here, we tested the association of single nucleotide variants with preeclampsia in 17,150 cases and 450,341 controls and with gestational hypertension in 8,961 cases and 184,925 controls from the InterPregGen study, FinnGen, Estonian Biobank, Michigan Genomics Initiative, BioMe, Genes & Health, Biobank Japan, and Mass General Brigham Biobank using multi-ancestry meta-analysis. We identified 11 genome-wide significant associations with preeclampsia, including 6 novel associations (*MTHFR* [1p36], *ARF1* [1q42], *HLA-S*/*MICA* [6p21], *TRPC6* [11q22], *FLT1* [13q12], *RGL3* [19p13]). In addition, we identified 4 novel associations with gestational hypertension, including two overlapping with preeclampsia (*FGF5* [4q21] and *ZNF831* [20q13]) and two unique associations (*NPR3* [5p13] and *CSAD* [12q13]). Colocalization and the Polygenic Priority Score method implicated several additional genes, e.g., *NPPA* and *CLCN6* (natriuretic peptide signaling), *SMARCA4* (trophoblast maintenance), *EPOR* (hypoxia response), and *KANK2* (podocyte function). A genome-wide preeclampsia polygenic risk score (PRS) predicted preeclampsia in the UK Biobank (OR 2.00 [95% CI 1.24-3.23], P=0.004, for top vs. bottom PRS decile); linear combination of the preeclampsia PRS with a multi-ancestry PRS for systolic blood pressure (SBP) improved preeclampsia prediction overall (OR 2.60 [95% CI 1.62-4.16], P<0.001, for top vs. bottom PRS decile), especially in non-European-ancestry individuals. Similarly, compared with a gestational hypertension PRS alone (OR 2.84 [95% CI 1.83-4.42], P<0.001, for top vs. bottom PRS decile), linear combination of gestational hypertension PRS with SBP PRS improved prediction of gestational hypertension (OR 4.23 [95% CI 2.54-7.06], P<0.001, for top vs. bottom PRS decile). Sex-stratified analyses in the UK Biobank revealed strong associations in women and men of the preeclampsia PRS with chronic hypertension (OR_{women}=1.15 per standard deviation [SD]; OR_{men}=1.12 per SD; $P_{interaction}$=$8.9x10^{-9}$) and ischemic heart disease ($P_{interaction}$>0.05) and of the gestational hypertension PRS with chronic hypertension (OR_{women}=1.15 per SD; OR_{men}=1.13 per SD; $P_{interaction}$=0.01) and ischemic heart disease ($P_{interaction}$>0.05). These findings provide mechanistic insights into the hypertensive disorders of pregnancy and advance pregnancy risk stratification.

S75. Gee, What A Session! (GWAS)

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 498. Genome-wide association study of chronic kidney disease progression in the Million Veteran Program and BioVU

Authors:

C. Robinson-Cohen[1], B. Rowan[1], J. He[1], O. Wilson[2], W. Wei-Qi[1], A. Bick[3], J. Hellwege[1], P. TSAO[4], J. Gaziano[5], M. Matheny[1], L. lipworth[1], C. Chung[1], E. Siew[1], T. Ikizler[1], R. Tao[1], Million Veteran Program, A. Hung[1,2]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Nashville VA, Tennessee Valley Hlth.care System, Nashville, TN, [3]Vanderbilt Univ., Nashville, TN, [4]VAPAHCS/Stanford Univ., Palo Alto, CA, [5]VA Boston Hlth.care System, Roxbury Crossing, MA

Abstract:

Rapid chronic kidney disease (CKD) progression is associated with poor clinical outcomes. Although the genetics of cross-sectional estimated glomerular filtration rate (eGFR) have been extensively studied, only a few loci associated with fast CKD progression, as assessed by eGFR decline over time, have been identified. We performed a meta-analysis of genome-wide association studies of eGFR decline among individuals with established CKD (stages 3 &4) in MVP (n=109,570) and BioVU (n=7,300). Our primary outcome was the annualized relative slope defined as longitudinal change in outpatient eGFR (% change per year), adjusted for age, sex, and 10 principal components of ancestry. Analyses were stratified by race and inverse variance-weighted fixed-effects meta-analyses were carried out using METAL. Analyses are stratified and presented by diabetes. Among participants with diabetes in the MVP and BioVU, mean (SD) eGFR at baseline was 52 (±7) ml/min and 49 (±17) ml/min, respectively, and median (IQR) relative kidney function decline was -1.3 (-4.2, 0.9) %/year and -2.38 (-10.8, +3.9) %/year respectively. Among participants without diabetes in the MVP and BioVU, mean (SD) eGFR at baseline was 52 (±8) ml/min and 51.4 (±18) ml/min, respectively, and median (IQR) relative kidney function decline was -0.40 (-2.1, 1.1) %/year and - 0.74 (-6.40, +4.41) %/year respectively. In the transethnic meta-analysis the strongest association in patients with diabetes was with rs77924615 near *PDILT*; each additional copy of the G allele was associated with a 0.45%/year faster decline in eGFR (p=2.24x10-13). Among participants without diabetes, the top association was the *UMOD/PDILT* variant rs36060036, associated with a 0.27%/year faster decline in eGFR per every copy of the C allele (p = 1.90 x10-17). A novel association was observed with *BICC1* (rs11592748); for every additional minor allele of rs11592748 there was a 0.14%/year slower decline in eGFR among individuals without diabetes (p = 6.73x10-9). Lead SNP rs11624911 near *HEATR4* was also associated, every additional copy of the A allele was associated with a 2.94%/year slower decline in eGFR (p=1.30x10-8) in African Americans. We also nominally replicated loci with known association with longitudinal eGFR changes (p<0.05) near candidate genes *TCF7L2* in diabetics and *PRKAG2, LINC009023, WDR72* and *APOL1* in individuals without diabetes. Three loci (two novel) were associated with longitudinal changes of eGFR at genome wide significance. We also replicated 5 candidate genes known to be associated with eGFR decline. Some of the novel identified loci provide potential for therapeutics.

S75. Gee, What A Session! (GWAS)

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 499. Trans-ancestry genome wide association of uterine fibroids in All of Us Research Program

Authors:

A. Williams[1], J. M. Keaton[1], H. Mo[1], D. J. Schlueter[1], C. Zeng[1], S. B. Goleva[1], T. Ferrara[1], O. Stubblefield[1], J. Dai[1], T. Cassini[2], J. Denny[1]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]NIH, Chevy Chase, MD

Abstract:

Despite the high prevalence of uterine fibroids and strong racial disparities between European Americans (EA) and African Americans (AA), electronic health records (EHR) have historically lacked racial and ethnic diversity for genetic analysis. Uterine fibroids are the most common benign pelvic tumors in females, presenting in at least 37% of hysterectomies in the United States (US). By age 50, AA have greater than 80% fibroid incidence compared to nearly 70% for EA. Racial disparities in onset age, number, size, and lifetime fibroid incidence by menopause also strongly support a role for genetic factors. We performed a trans-ancestry GWAS of uterine fibroid risk in women of predominantly African and European ancestries in *All of Us* Research Program. The goal was to identify unique genetic variants associated with fibroid risk across multiple ancestry groups. All of Us is longitudinal cohort study that aspires to enroll more than one million participants representing diverse populations in the US and 80% includes those historically underrepresented in biomedical research. For this study, we included 15706 participants with both whole genome sequencing (WGS) data and EHR data available. Cases were defined as females over age 18 with 1 fibroid diagnostic code confirmed with pelvic imaging, 2 or more fibroid diagnostic codes, or self-reported fibroids on the Personal Medical History (PMH) survey (N=6154). Controls were defined as females over age 18 with 2 pelvic imaging procedure codes and no fibroid diagnostic codes or indication of fibroids in the PMH survey (N=9552). These data included 2849 AA (1243 cases,1606 controls) and 7912 EA (3522 cases,4390 controls). GWAS was performed using logistic regression adjusted for age and the first seven principal components. We identified 2 genome-wide significant loci including the known locus *THEMP7/WT1* (rs73309006, OR=1.66 ± 0.12, p=1.39 E-09) and a novel locus at *ZNF831* (rs116540272, OR= 3.55 ± 0.70, p=4.02 E-08). Rs73309006 is in strong linkage disequilibrium (r^2 &gt 0.8) with 2 regulatory SNPs, rs80225530 and rs74316501 which likely affect transcription factor binding. *ZNF831* has been associated with hypertension outcomes in previous GWAS. We also detected nominal association at *MMP26* (index rs80191705, OR= 3.80 ± 0.82, P=1.38E-07) which encodes matrix metalloproteinase 26 and is involved in the breakdown of extracellular matrix in reproduction and tissue remodeling. These results demonstrate the potential of multi-ancestry studies in identifying genetic risk factors for fibroids and highlights possible pleiotropic effects of genomic risk factors for hypertension on fibroid risk.

S76. Genetic architecture of adiposity

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 502. Characterising the circulating proteome of adiposity through use of weight loss interventions and Mendelian randomisation

Authors:

L. Goudswaard[1], M. L. Smith[1], D. Hughes[1], N. Sattar[2], R. Taylor[3], M. E. J. Lean[2], J. M. Blazeby[1], NIHR By-Band-Sleeve Trial Management Group, I. Hers[1], L. J. Corbin[1], N. J. Timpson[1]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]Univ. of Glasgow, Glasgow, United Kingdom, [3]Newcastle Magnetic Resonance Ctr., Newcastle, United Kingdom

Abstract:

Adiposity is associated with an increased risk of type 2 diabetes (T2D) and cardiovascular disease. It is likely that a change in circulating proteins plays a role in obesity-related disease risk. This study aimed to triangulate evidence from randomized controlled trials of caloric restriction and bariatric surgery, along with a Mendelian randomisation (MR) study to characterise the effects of BMI on proteins in circulation. Data were used from the intervention arm of the Diabetes Remission Clinical Trial (DiRECT, N=119). Here, participants with overweight/obesity (baseline mean BMI 35.0 kg/m^2, SD 4.6 kg/m^2) underwent the CounterWeight Plus weight management programme, consisting of a low energy total diet replacement (TDR). Plasma samples were taken at baseline and after 1-year and SomaLogic was used to quantify plasma proteins. Data were also utilised from a pilot sample release as part of an ongoing trial of surgically induced weight loss (N=119). Serum samples were donated at baseline and 3-years post bariatric surgery; serum proteins were measured by the Olink Explore 1536 panel. Protein outcomes were inverse rank normal transformed and linear mixed models used to determine proteins associated with the weight loss interventions (3 types of bariatric surgery). Proteins with consistent evidence across both trials were looked up in our previously published MR analysis, which used a UK blood donor cohort, INTERVAL (N=2729), to estimate the causal effect of BMI on circulating plasma proteins. 1496 proteins out of 4601 were altered with the TDR intervention ($p<5.8x10^{-5}$) and 188 proteins out of 1472 were associated with bariatric surgery ($p<6.2x10^{-5}$). 40 unique proteins were altered in the same direction with both interventions, of which 11 also had consistent MR effect estimates. These include intervention-associated increases in levels of insulin-like growth factor binding proteins 1/2 and ecto-ADP-ribosyltransferase 3, with higher BMI being associated with lower levels in an MR framework. Interventions reduced levels of alcohol dehydrogenase 4, glutathione S-transferase A1 and mannan-binding lectin serine protease 1, with higher BMI being associated with higher levels in an MR framework. This study utilized three independent cohorts and distinct analytical approaches to explore the effect of BMI on circulating proteins. Observing consistency across such analyses provides greater confidence that the effects we see reflect genuine physiological responses to differential BMI. We also observe discordance between association results that hint at intervention specific signals; these will form the focus of future work.

S76. Genetic architecture of adiposity

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 503. Identification of rare functional variants in known genes for monogenic obesity among American Indians with severe obesity

Authors:

C. Koroglu Altok[1], Y. L. Muller[1], M. Traurig[1], S. Day[1], P. Piaggi[1], K. Wiedrich[1], R. L. Hanson[1], W. C. Knowler[1], Regeneron Genetics Center, C. V. Van Hout[2], A. R. Shuldiner[2], C. Bogardus[1], L. J. Baier[1]; [1]NIDDK, NIH, Phoenix, AZ, [2]Regeneron Genetics Ctr., Tarrytown, NY

Abstract:

Southwestern American Indians (SWAIs) have a high rate of obesity. This study aimed to identify variants in previously reported monogenic obesity genes that potentially contribute to the severe cases of obesity among SWAIs. Anticipating that these variants would be extremely rare and thus lack statistical power in a genome-wide association analysis, we adopted an approach whereby only variants most likely to affect protein function were 1) analyzed for their occurrence in severely affected cases versus controls, and 2) functionally studied *in vitro*. Analysis of whole-exome sequencing data acquired from 6,803 longitudinally studied SWAIs identified 301 missense and splicing variants located in 15 genes previously shown to cause monogenic obesity. These 301 variants were evaluated for a potential role in extreme obesity by analyzing their occurrence in 2 separate case and control samples consisting of children (&lt 18 yrs) or adults (&gt 18 yrs). These analyses identified 8 missense variants present only in individuals with severe obesity. Four variants were detected in a child(ren) with a maximum age-sex- and population-specific BMI z-score &gt 2 (N = 279), but not found in any child with a maximum BMI z-score &lt 0 (N = 1,542). For reference, in this population BMI z-scores 2 and 0 are the equivalent of BMI 41 kg/m^2 and 27 kg/m^2, respectively in a 14-year-old female. Five variants (one overlapped with childhood analysis) were detected among the adults in the top 5th percentile of BMI in this sample (N = 271; BMI &gt 53.13 kg/m^2) but not in adults below the median BMI (N = 2,727; BMI &lt 35.92 kg/m^2). The 8 missense variants are located within 6 monogenetic obesity genes: *KSR2*, *MC4R*, *DYRK1B*, *NTRK2*, *PCSK1*, and *SIM1*, the latter four genes each carry a single mutation observed in a single severe obesity case. *KSR2* had two novel variants (p.I402F and p.T193I) carried by three cases. Two variants in *MC4R* (p.A303P and p.R165G), which together were found in 2.5% of childhood cases, were previously reported by our group to decrease MC4R function and therefore were not further studied. The remaining six variants were studied *in vitro* using reporter assays (*KSR2*, p.I402F and p.T193I; *NTRK2*, p.S249Y; *SIM1*, p.R383K; and *DYRK1B*, p.P495L) or an ELISA (*PCSK1*, p.R740Q). We found the p.S249Y in NTRK2, and both the p.I402F and p.T193I in KSR2, appear to alter ERK signaling. The p.R383K in SIM1 downregulated luciferase expression via a *JAK2* promoter reporter vector. No effect was observed with the p.P495L in DYRK1B, while results of the p.R740Q in PCSK1 are pending. In summary, several rare missense variations in known monogenic obesity genes may contribute to the most severe obesity in American Indians.

S76. Genetic architecture of adiposity

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 504. Discovery of novel loci that uncouple adiposity from its cardiometabolic comorbidities in the UK Biobank: towards precision medicine of obesity

Authors:

N. Chami[1,2], Z. Wang[1,2], D. Hemerich[1], R. J. F. Loos[1,2,3]; [1]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]The Mindich Child Hlth.and Dev. Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]The Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark

Abstract:

Individuals with obesity are at a higher risk of developing cardiometabolic outcomes such as T2D and CVD. However, some individuals with obesity do not suffer from these conditions. We and others have identified dozens of variants that contribute to the uncoupling of obesity from cardiometabolic co-morbidities using GWAS summary statistics. Here, we leveraged individual-level data of the UK Biobank and performed GWAS analyses on 35 adiposity and cardiometabolic risk factors and identified 266 variants associated with increased adiposity but with protective cardiometabolic effects. We replicate most of the previously known loci (49 loci), including *IRS1*, *PPARG, RSPO3, VEGFA*, and *KLF14*. In a PheWAS analysis, a GRS of the 266 variants (GRSprotective) was associated with increased risk of obesity (OR=1.16, $P<4x10^{-108}$), but a decreased risk of several cardiometabolic outcomes, including disorders of lipoprotein metabolism (OR=0.92, $P= 1.41x10^{-89}$), non-insulin-dependent diabetes (OR=0.94, $P= 5.56x10^{-21}$), ischemic heart disease (OR=0.96, $P=7.38x10^{-11}$), and essential primary hypertension (OR=0.94, $P=1.68x10^{-27}$). The 266 loci were validated in BioMe ($N_{samples}$=23k); individuals in the top decile of the GRSprotective had higher BMI (mean=29.6 vs 28.0 kg/m^2, $P=1.2x10^{-12}$), but lower triglycerides (mean=130.5 vs 148.7 mg/dl, $P=1.6x10^{-6}$), LDL (mean= 111.5 vs 115.6 mg/dl, P=0.02), and HbA1C (6.15 vs 6.29%, P=0.03), and higher HDL (mean=55.7 vs 53.9 mg/dl, P=0.007) compared to individuals in the lowest decile. Similar to the UK Biobank, GRSprotective was also associated with higher obesity (OR=1.08, $P=1.6x10^{-4}$) and lower CVD outcomes such as mitral valve disease (OR=0.72, $P=5.1x10^{-4}$) and congestive heart failure (OR=0.86, $P=8.2x10^{-4}$). Clustering analyses grouped the 266 variants into 8 clusters, each with a distinct signature, revealing subsets of variants that act differentially on adiposity and one or more cardiometabolic risk factor. Three clusters showed protective effects on multiple cardiometabolic risk factors and were associated with lower liver fat (Beta=-0.19, P=3.6x10-6; Beta=-0.09, P= 0.004; Beta=-0.07, P=0.003). Unlike genes located in body fat percentage (BFP) associated loci which were mainly enriched in the nervous system- consistent with previous findings- genes within the 266 loci were not enriched in the nervous system and were highly expressed in adipose tissue, and in tissues of the cardiovascular, endocrine, and muscoskeletal systems (FDR<0.05). In summary, we identified 266 variants that uncouple adiposity from cardiometabolic outcomes and cluster into 8 subsets with distinct cardiometabolic risk profiles.

S76. Genetic architecture of adiposity

Location: Conv Ctr/Room 515/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 505. Rare loss of function variants in the hepatokine gene *INHBE* protect from abdominal obesity

Authors:

A. Deaton[1], A. Dubey[1], L. D. Ward[1], P. Dornbos[2], J. Flannick[3], AMP-T2D-GENES Consortium, E. Yee[1], S. Ticau[1], L. Noetzli[1], M. M. Parker[1], R. Hoffing[1], C. Willis[1], M. E. Plekan[1], A. M. Holleman[1], G. Hinkle[1], K. Fitzgerald[1], A. K. Vaishnaw[1], P. Nioi[1]; [1]Alnylam Pharmaceuticals, Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Boston Children's Hosp., Boston, MA

Abstract:

Identifying genetic variants associated with lower waist-to-hip ratio can reveal new therapeutic targets for abdominal obesity and metabolic syndrome. We used whole exome sequencing data from 362,679 individuals in the UK Biobank to identify genes associated with waist-to-hip ratio adjusted for BMI (WHRadjBMI), a surrogate for abdominal fat that is causally linked to type 2 diabetes and coronary heart disease. Rare predicted loss of function (pLOF) variants in *INHBE*, carried by 1 in 587 individuals, were associated with lower WHRadjBMI ($P = 5 \times 10^{-8}$, 0.22 standard deviation decrease) and this association replicated in data from the AMP-T2D-GENES consortium. *INHBE* encodes a secreted protein, the hepatokine activin E. In vitro characterization of the most common *INHBE* pLOF variant in our study, a splice acceptor variant, indicates an in-frame deletion resulting in an approximately 90% reduction in secreted protein levels. We also detected associations with lower WHRadjBMI for variants in *ACVR1C*, encoding an activin receptor, further highlighting the involvement of activins in regulating fat distribution. We identified associations for 10 additional genes in our primary analysis, including the Mendelian lipodystrophy gene *PLIN1*, as well as four sex-specific associations. These findings highlight activin E as a potential therapeutic target for abdominal obesity, a phenotype causally linked to cardiometabolic disease.

S77. Genetic variants and cancer risk in diverse populations

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 508. Risk allele associated with childhood acute lymphoblastic leukemia at the IKZF1 locus is associated with Indigenous American ancestry and absent in European ancestry populations

Authors:

S. Jeon[1], S. Gazal[1], N. Mancuso[1], C. Metayer[2], X. Ma[3], J. Wiemels[1], C. Chiang[1], A. de Smith[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ. of California, Berkeley, Berkeley, CA, [3]Yale Sch. of Publ. Hlth., New Haven, CT

Abstract:

Children of Hispanic/Latino ethnicity have the highest risk of acute lymphoblastic leukemia (ALL) in the US, with up to 40% increased incidence than in non-Latino whites. Differences in the presence or frequencies of risk alleles may partly explain such disparities in disease risk. These population-enriched risk alleles may be missed due to lack of power in genetic association studies if they are more prevalent in a minority population. Further, these alleles could be secondary but independent associations at a functional locus, and so their effect on disease risk is further masked. Here, using the largest-available multi-ancestry genome-wide association study (GWAS) for childhood ALL, we identify population-specific secondary signals at the IKZF1 gene. Prior European-centric or multi-ancestry GWAS identified two independent associations at IKZF1. Using genome-wide SNP array data from 1878 childhood ALL cases and 8441 controls of self-reported Latino ethnicity in California, imputed with the TOPMedImputation server, we identified a third independently associated variant in this locus (P = 4.67x10-11, OR = 1.44). The newly identified risk allele is nearly monomorphic in non-Finnish Europeans (frequency = 0.2%), but is enriched in Latinos (22.3%) and East Asians (20.3%) in gnomAD. Hence, this risk locus was not identified in a GWAS of ALL including 1162 cases and 57,341 controls of self-reported non-Latino white race/ethnicity, in which only two independent IKZF1 risk loci were detected. Among Latinos, the new risk allele is significantly associated with both global and local Indigenous American ancestry; in fact, the IKZF1locus would achieve genome-wide significance through admixture mapping (P = 8.78x10-7) in a relatively small sample of 4032 Latino individuals, suggesting that admixture mapping could be fruitful in identifying additional population-specific risk loci. Two of the three independent IKZF1 risk loci in Latinos, including the new third risk allele, are tightly linked to nearby SNPs downstream of IKZF1 that are 26bp apart and within the same enhancer locus predicted to target IKZF1 in hematopoietic cell types including CD34 cells and B-cells in EpiMap. Analysis is ongoing to elucidate the impact of these risk alleles on the activity of this IKZF1 enhancer.In sum, we observed three independent IKZF1 risk loci for childhood ALL among Latinos, while only two loci are found in non-Latino whites. The third locus is only apparent when imputed with TOPMed due to its increased representation of IndigenousAmerican haplotypes. Admixture mapping may also have detected this locus, and others, that contribute to the increased risk of ALL in Latinos.

S77. Genetic variants and cancer risk in diverse populations

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 509. A multi-population African GWAS for prostate cancer reveals novel disease associations and within-continent heterogeneity of cancer risk

Authors:

R. Janivara[1], U. Hazra[1], C. Chen[2], J. Lachance[1], T. Rebbeck[3], MADCaP (Men of African Descent and Carcinoma of the Prostate) Network; [1]Georgia Inst. of Technology, Atlanta, GA, [2]Univ. of the Witwatersrand, Johannesburg, South Africa, [3]Dana Farber Cancer Inst., Boston, MA

Abstract:

Men of African descent have elevated incidence and mortality rates of prostate cancer (CaP), yet its genetic basis in Africa has historically been understudied. This knowledge gap is exacerbated by the genetic heterogeneity found within the continent. Here, we present a trans-African CaP GWAS from the Men of African Descent and Carcinoma of the Prostate (MADCaP) Network, which is the first of its kind. It provides a unique opportunity to characterize the genetic architecture of CaP in African populations, fine-map disease associations, and help remedy disparities in genomic medicine. A total of 4,728 samples from seven study sites across Ghana, Nigeria, Senegal, and South Africa were analyzed using data from our specialized MADCaP array. All the samples were genotyped in Africa, and cases and controls were ancestry-matched for each study site. Over 15 million imputed variants were tested for genetic associations between case and control status and disease severity. Linear-mixed models were used to find associations separately for Western and Southern Africa, and a meta-analysis was conducted for the combined dataset. Our findings include replication and fine-mapping of associations originally found in non-African populations, the discovery of novel African associations, as well as evidence of within-continent heterogeneity of CaP risk. Significant associations were found in the 6q22 region near RFX6. Complex associations with CaP were found in the 8q24 region near PCAT1. Intriguingly, the lead SNPs in this region (rs72725854 and rs73705708) have negligible allele frequencies outside of Africa, emphasizing the utility of GWAS studies on diverse human populations. Geographically specific associations were observed for multiple genomic regions (including 10q26 and 17q24). Evolutionary genetic analyses identified CaP-associated variants with large allele frequency differences between Western and Southern Africa at 9p24 and 6p21. These allele frequency differences, as well as potential genotype-by-environment interactions, contribute to within-Africa heterogeneity in the risk of CaP. Leveraging information on disease aggressiveness (age of onset, tumor grade, and Gleason score), we also generated a novel African CaP polygenic risk score and validated its effectiveness using external datasets. This work advances our understanding of the genetic etiology of a complex disease that has an uneven burden across the globe and the African continent.

S77. Genetic variants and cancer risk in diverse populations

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 510. MeSuSie: a novel method for discovering shared and unique putative causal variants by fine-mapping across diverse ancestries

Authors:

B. Gao, X. Zhou; Univ. of Michigan, Ann Arbor, MI

Abstract:

Understanding ancestry specific genetic architecture of complex traits is a crucial question in GWAS. Joint fine-mapping leverages information across ancestries and improves accuracy and resolution in shared causal variant detection. However, previous approaches for jointly fine-mapping signals from multiple ancestries have limited the combinations of causal status by assuming all causal variants are shared across ancestries; the possibility of ancestry-specific variants is ignored and hence causes false discovery of shared variants. Here we develop a method with a more general assumption on the causal status of variants across ancestries that bridges this gap. We use a mixture of the multivariate normal distribution to model the effect of causal variants, which can capture both ancestry-specific and shared causal variants. We refer to our method as the multi-ancestry sum of the single effect model (MeSuSie). We perform extensive simulations including the baseline setting where 50% of causal variants are shared across ancestries and the alternative setting where all causal variants are shared. We further examine the impact of the sample size of GWAS, effect size correlation of shared signals, and LD matrix accuracy to evaluate the robustness of MeSuSie. We compare the performance of MeSuSie with ancestry-specific SuSie and the multi-ancestry fine-mapping method Paintor. At common PIP thresholds (0.9, 0.5) in both baseline and alternative simulation settings, our proposed method has increased power over SuSie and better control of FDR over Paintor, and is robust to inaccurate LD structure and a wide range of parameter settings. We further apply MeSuSie for fine mapping of four lipid traits (HDL, LDL, TC, and TG) using GWAS summary statistics from European samples in UKBB and from African samples in GLGC in the 798 peak GWAS variant harboring region. The proportion of ancestry-specific variants ranges from 57.1%-72.2%, which is consistent with previous findings. The manual inspection confirms that MeSuSie is more likely than competing methods to identify and correctly label putative shared or ancestry-specific signals.

S77. Genetic variants and cancer risk in diverse populations

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 511. Genotype-first breast cancer prevention - experience with transferring monogenic findings from a population biobank to clinical setting

Authors:

M. Nõukas[1], H. Jürgens[2], L. Roht[3,4], L. Leitsalu[5], M. Palover[6], T. Nikopensius[6], A. Reigo[6], M. Kals[6,7], K. Kallak[2,8], R. Kütner[9], K. Budrikas[10], S. Kuusk[6,10], V. Valvere[9], P. Laidre[11], K. Toome[11], K. Rekker[11], M. Tooming[4,11], Ü. Murumets[11], T. Kahre[4,11], K. Kruuv-Käo[6,12], K. Ounap[4,11], P. Padrik[2,8,13], A. Metspalu[6], T. Esko[6], K. Fischer[6,10], N. Tõnisson[4,11,6]; [1]Univ. of Tartu, Tartu, Estonia, [2]Tartu Univ. Hosp., Clinic of Hematology & Oncology, Tartu, Estonia, [3]Tartu Univ. Hosp., United Lab., Tallinn, Estonia, [4]Dept. of Clinical Genetics, Inst. of Clinical Med., Univ. of Tartu, Tartu, Estonia, [5]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [6]Estonian Biobank, Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [7]Inst. for Molecular Med. Finland (FIMM), HiLIFE, Univ. of Helsinki, Helsinki, Finland, [8]Univ. of Tartu, Clinic of Hematology & Oncology, Tartu, Estonia, [9]North-Estonian Med. Ctr., Oncology and Haematology Clinic, Tallinn, Estonia, [10]Univ. of Tartu, Inst. of Mathematics and Statistics, Tartu, Estonia, [11]Dept. of Clinical Genetics, United Lab., Tartu Univ. Hosp., Tartu, Estonia, [12]Tartu Univ. Hosp., Tartu, Estonia, [13]Antegenes, Tartu, Estonia

Abstract:

Background/Objectives: Although hereditary breast cancer (BC) screening and management is well established in clinical setting, it only captures a fraction of genetic predisposition at the population level. We have addressed this gap by returning data on BC-predisposing monogenic variants to female participants of the population-based cohort in the Estonian Biobank (EstBB) as a possible future service based on affordable genotyping data broadly available in personalized medicine era. **Methods:** We describe our experience from a national pilot study (2018-2021) in collaboration with two Estonian regional hospitals. We screened the EstBB cohort (n=154,201; 100,731 women [66.1%]) to identify female carriers aged 22-79 years with monogenic variants conferring high or moderate BC risk in any of 11 genes listed in clinical guidelines (*BRCA1*, *BRCA2*, *TP53*, *STK11*, *PTEN*, *CDH1*, *ATM*, *PALB2*, *CHEK2*, *NBN*, *NF1*). We confirmed BC-associated variants and re-contacted 180 female carriers. Total of 109 (61%) participants consented for further clinical assessment by clinical geneticists and oncologist, that included confirmation of genetic variant from a new blood sample, imaging studies, laboratory testing, return of results, preventive surgery and cascade screening. To verify the cumulative BC risks, all female participants and full list of validated monogenic pathogenic BC predisposing variants in the EstBB were included. **Results:** Our results show that only one-third of participants would have been eligible for BC screening according to the current clinical criteria showing the demand for a broader genetic screening for monogenic risk factors. Cancer was diagnosed in 16 participants prior to our study and only 3 of them were aware of their genetic risk information. BC was diagnosed in six participants. Majority of the participants considered receiving the genetic risk information as valuable. We also confirmed that the hazard ratios for *BRCA1*/2 variant carriers (12.1; 95% CI 9.1 to 16.0) and for *CHEK2* variant carriers (4.4; 95% CI 2.7 to 10.7) fully support genotype-based clinical interventions that could be further applied in population-based personalized prevention for hereditary BC. **Grants:** This research was funded by the Estonian Research Council-RITA programme, by the European Regional Development Fund in accordance with Directive no. 1.1-2/17/15; Project No. 2014-2020.4.01.15-0012 GenTransMed, and Estonian Research Council grants PRG555 and PRG1197.

S78. Let me sleep on it

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 514. Whole-exome sequencing study identified novel genes for self-reported, diagnosis, and accelerometer-based sleep and circadian rhythm traits and disorders

Authors:

C-Y. Chen[1], S. Jones[2], J. M. Lane[3], A. Wood[4], Biogen Biobank Team, R. Saxena[5], M. Weedon[6], H. M. Ollila[2]; [1]Biogen, Cambridge, MA, [2]Univ. of Helsinki, Helsinki, Finland, [3]Brigham and Women's Hosp., Boston, MA, [4]Coll. of Med. and Hlth., Exeter Univ., Exeter, United Kingdom, [5]Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [6]Univ. of Exeter, Exeter, United Kingdom

Abstract:

Genome-wide association studies (GWAS) have identified loci associated with sleep and circadian traits and disorders, which have adverse impacts on many health outcomes (e.g., cardiometabolic diseases). To further our understanding of the genetic basis of sleep and circadian traits and disorders, we perform systematic analyses on the impact of rare-coding and common variants on 46 self-reported, diagnosis-, and accelerometer-based sleep and circadian phenotypes in over 450,000 whole-exome sequenced and genotyped samples in UK Biobank (UKB). To expand on previous findings, we align the phenotype definitions to those in the previous GWAS. We first show significant impact of exome-wide rare protein truncating variants (PTV) and missense variant burden on sleep phenotypes across the board, with stronger impacts from variants with lower minor allele frequency (MAF) and in LoF intolerant genes. We identify genes associated with sleep phenotypes through PTV and missense burden tests, including previously reported *PER2* (PPTV=1.12×10^{-11}) and *MTNR1B* (Pmissense=1.51×10^{-13}) for morningness and *PER3* (PPTV=9.75×10^{-9}) and *RASEF* (PPTV=2.50×10^{-8}) for accelerometer-derived sleep timing and sleep duration variability, as well as novel genes such as *SERPINB6* (PPTV=4.61×10^{-10}) and *PLXNB2* (PPTV=1.10×10^{-8}) for sleep medication use and *ANKRD12* (PPTV=3.13×10^{-8}) for self-reported daytime sleepiness. Replication is performed with independent exome sequenced datasets including Mass General Brigham Biobank and All of Us Program. We then use exome data to fine-map GWAS associated signals, implicate causal genes at GWAS loci, and pinpoint causal variants with functional implications. For example, we identify 10 missense variants in MTNR1B in the UKB exome sequence, where the top missense variant associated with morningness rs61746674 (P=6.56×10^{-7}) is a rare variant (MAF=0.0015) previously associated with type 2 diabetes. Mendelian randomization and colocalization with brain and blood eQTLs and pQTLs are performed to provide further functional insights. We further compare the rare coding and common variation heritability and genetic correlations across sleep phenotypes within UKB and with previous GWAS and examine the relative variance explained by common variant polygenic scores and rare coding variant burden across sleep phenotypes. In summary, our study identifies novel genes and provides a comprehensive view of the rare-coding and common genetic architecture of sleep and circadian phenotypes and disorders that can reveal underlying biological processes and pleiotropic links to disease, and serve as a foundation for novel therapeutics developments.

S78. Let me sleep on it

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 515. Enabling PheWAS for psychiatric disorders through deep phenotyping of behavior from smartphones sensors

Authors:

B. Balliu, C. Douglas, L. Shenhav, A. Wu, N. Freimer, E. Halperin, S. Sankararaman, J. Flint; Univ. of California, Los Angeles, Los Angeles, CA

Abstract:

The explosion of GWAS discoveries from the UK Biobank has fostered entirely new paradigms for genetic discovery, e.g., the phenome-wide association study (PheWAS), the systematic search of phenotype databases for measures associated with a given variant. So far, however, this transformation has largely bypassed psychiatry, which currently lacks large cohorts phenotyped for quantitative traits. This lack reflects primarily the difficulty in obtaining meaningful quantitative assays of human mood and behavior that are both scalable to genetic studies and relevant to clinical disorders.Here, we use computerized adaptive testing (CAT) to obtain high-quality longitudinal assessments of depressed mood on 180 individuals with varying degrees of depressive symptoms over 10 months. In addition, we use sensors from smartphones to perform deep digital behavioral phenotypes passively and continuously. In total, we obtained 3,005 days with mental health assessments and 29,254 days with behavioral phenotype assessments. We then applied idiographic (personalized) and nomothetic supervised models to the multidimensional data to generate individualized predictions of depressed mood up to six weeks in advance.We show that idiographic prediction models are superior to nomothetic models. The nomothetic model yielded a mean absolute percent error of 22% while the ideographic models had a mean absolute percent error of 17. In addition, using an idiographic model, we were able to accurately predict mood up to four weeks in advance (R-squared= 78.9%). In order to obtain behavioral features that can then be used for downstream PheWAS of mental health, we extracted top-feature predictors. We found that behavioral features related to sociability, e.g., percentage of outgoing calls, and sleep disruption, e.g., time spent at home at night, contributed most to prediction accuracy.In conclusion, our study verified the feasibility of using passively collected deep digital behavioral phenotypes from smartphones to predict depressive symptoms weeks in advance. Its key novelty lies in the use of computerized adaptive testing, which enabled us to obtain high-quality longitudinal assessments of mood on hundreds of individuals over many months, and in the use of idiographic prediction models, which offer a much higher predictive power compared to nomothetic models. Ultimately, we expect that the method will lead to large GWAS and PheWAS of psychiatric disorders.

S78. Let me sleep on it

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 516. Identification of a putatively adaptive promoter variant in *PRKAA1* in Andean highlanders and associations with hypoxia adaptation

Authors:

S. Fassardi[1], E. S. Lawrence[1], W. Gu[1], M. C. Kong[1,2], A. Lin[1], E. A. Moya[1], J. E. Hall[1], K. A. O'Brien[1,3], T. S. Simonson[1]; [1]Dept. of Med., Div. of Pulmonary, Critical Care, and Sleep Med., Univ. of California, San Diego, La Jolla, CA, [2]Dept. of Bioengineering, Univ. of California, San Diego, San Diego, La Jolla, CA, United States of America, La Jolla, CA, [3]Dept. of Physiology, Dev. and NeuroSci., Univ. of Cambridge, Cambridge, CB2 3EG,, Cambridge, United Kingdom

Abstract:

Andeans have resided at high altitude for thousands of years and exhibit distinct physiological and metabolic phenotypes to maintain homeostasis under the selective pressure of hypoxia. Based on previous whole-genome sequence data generated in 40 Andeans from Cerro De Pasco Peru (4340m), we utilized the composite of multiple signals test of selection and identified an adaptive signal at *PRKAA1*, which encodes the alpha-1 subunit of the AMP-activated protein kinase (AMPK). *PRKAA1* was one of the top scoring genes in our selection scan results (CMS Bayes Factor = 14.31) and has been previously reported to contribute to the hypoxic ventilatory response in mice. We hypothesize variants in *PRKAA1* are associated with ventilatory control in highland populations exposed to chronic hypoxia. While no protein-coding variants were identified as potential functional targets of selection, we examined variants within regions of known epigenetic activity near *PRKAA1*. The variants identified were in high linkage-disequilibrium (LD) with our selection scan markers as well as previously reported *PRKAA1* variants associated with uterine artery diameter and other fetal phenotypes in Andeans. We demonstrated associations between a *PRKAA1* promoter variant rs10035235 (chr5:40798644, C>T) and control of breathing phenotypes in Andeans. Males with additional copies of the putatively adaptive T allele were shown to have increased hypoxic ventilatory response ($p < 0.03$). The variant was also associated with end-tidal CO2 ($p < 0.01$), ventilation ($p < 0.05$), and heart rate ($p < 0.05$). We further examined the effect of rs10035235 on *PRKAA1* gene expression using a luciferase assay in HEK293 cells and found the T allele resulted in a 203% increase in expression relative to cells with the non-adaptive C allele ($p < 0.05$).
To assess whether rs10035235 was associated with other phenotypes related to control of breathing, we tested for associations with a subset of phenotypes from a large cohort comprised of Hispanics and Latinos (HCHS-SOL) (n = 11,893). We identified associations with apneic events (FDR < 3.30E-3), total time spent in apnea (FDR < 7.90E-4) as well as the apnea/hypopnea index (FDR < 3.09E-3). These findings suggest adaptation at the *PRKAA1* locus is associated with increased breathing responses to hypoxia in Andean highlanders and may underlie variation in sleep-apnea related phenotypes in the Hispanic/Latino population.

S78. Let me sleep on it

Location: Conv Ctr/Room 502/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 517. An integrated multi-omics analysis of sleep disordered breathing traits across multiple blood cell types

Authors:

N. Kurniansyah[1], D. Wallace[1], Y. Zhang[1], B. Yu[2], B. Cade[3], H. Wang[4], H. Ochs-Balcom[5], A. Reiner[6], A. Ramos[7], J. Smith[6], J. Cai[8], P. Zee[9], R. Kaplan[10], C. Kooperberg[11], S. Rich[12], J. Rotter[13], S. Gharib[6], S. Redline[1], **T. Sofer**[1]; [1]Brigham and Women's Hosp., Boston, MA, [2]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Brigham & Women's Hosp. / Harvard Med. Sch., Boston, MA, [4]Brigham and Women's Hosp., Boston, MA, [5]Dept. of Epidemiology and Environmental Hlth., Sch. of Publ. Hlth.and Hlth.Professions, Univ. at Buffalo, Buffalo, NY, [6]Univ. of Washington, Seattle, WA, [7]Univ. of Miami Miller Sch. of Med., Miami, FL, [8]Univ. of North Carolina, Chapel Hill, NC, [9]Northwestern Univ., Chicago, IL, [10]Albert Einstein Coll. of Med., Bronx, NY, [11]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [12]Univ. of Virginia, Scottsville, VA, [13]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA

Abstract:

Background: Sleep Disordered Breathing (SDB) is characterized by repeated breathing reductions or cessations during sleep, often accompanied by oxyhemoglobin desaturation. How SDB affects the molecular environment is still poorly understood.
Methods: We studied the association of three SDB measures: the Apnea Hypopnea Index (AHI), average and minimum oxyhemoglobin saturation during sleep (AvgO2 and MinO2) with gene expression measured using RNA-seq in peripheral blood mononuclear cells (PBMCs), monocytes, and T-cells, in ~500 individuals from the Multi-Ethnic Study of Atherosclerosis (MESA). We developed genetic instrumental variables (IVs) for the associated transcripts as polygenic risk scores (tPRS), then generalized and validated the tPRS in the Women's Health Initiative (WHI). Next, we constructed the tPRS and studied their association with objective measures of SDB (to identify potential reverse causal associations) and with plasma metabolites (to identify downstream effects) in ~12,000 and ~4,000 participants, respectively, from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Finally, we estimated the association of these SDB measures with transcript IV-associated metabolites in HCHS/SOL, to verify complete association pathways linking SDB, gene expression, and metabolites.
Results: Across the three leukocyte cell types, 96 gene transcripts were associated with at least one SDB exposure (False Discovery Rate (FDR) <0.1). Across cell populations, estimated log-fold expression changes were similar between AHI and MinO2 (Spearman correlations>0.90), and less similar between AvgO2 and the other exposures. Eight and four associations had FDR < 0.05 when the analysis was not adjusted and adjusted for BMI, respectively. Associations include known genes that respond to hypoxia (*PDGFC*) and regulate this response (*AJUBA*). We identified a complete "chain" linking AvgO2, *P2RX4*, and the metabolite butyrylcarnitine (C4), suggesting that increased expression of the purinergic receptor *P2RX4* may be associated with improved average oxyhemoglobin saturation and decreased butyrylcarnitine (C4) levels.
Conclusions: Our results support a mechanistic role for purinergic and hypoxic signaling, among others, in SDB. These findings show differential gene expression by blood cell type in relation to SDB traits and link *P2XR4*expression to influencing AvgO2 and butyrylcarnitine (C4) levels. Overall, we employed novel methods for integrating multi-omic data to evaluate biological mechanisms underlying several SDB traits.

S79. Splicing together the story

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 520. Single-cell differential splicing of Alzheimer's disease across 1.9 million cells and 416 individuals

Authors:

C. Hwa[1], L. He[2], Y. Park[3], L-L. Ho[1], H. Mathys[4], L-H. Tsai[1], M. Kellis[1]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Duke Univ., Durham, NC, [3]Univ. of British Columbia, Vancouver, BC, Canada, [4]Univ. of Pittsburgh, Pittsburgh, PA

Abstract:

Alzheimer's Disease (AD) is a polygenic disease with variable phenotypic response among patients. While single-cell differential expression data can provide important information regarding the disease's pathology, it only captures steady-state cell behavior. By contrast, RNA velocity analysis allows inference of both direction and magnitude of regulatory changes occuring over disease progression. However, despite the success of existing RNA velocity analyses, the pace of disease progression remains uncharacterized, as previous work focused primarily on macro-scale trajectory diagrams and failed to capture trajectories of individual genes relative to global pathological changes. Here, we use scRNA-seq profiling of 1.9 million cells from dorsolateral prefrontal cortex of 416post-mortem human samples to model temporal regulatory changes of AD. We use RNA velocity to infer spliced and unspliced counts for each gene and a beta-binomial distribution to model the unspliced/spliced ratio, and we identify differentially-spliced genes across six brain cell types and differentially-spliced genes between AD and control cohorts. Of 16,948expressed genes, 1158 genes show significant differential splicing (DS) effects with FDR<0.05%. These include 38 previously-known AD GWAS genes, and 1120additional DS genes, greatly expanding previous candidates, increasing classification accuracy, and revealing the dynamics of these genes across early-, middle-, and late-stage AD pathology progression. We find that cell-type specific and globally-altered genes are significantly enriched in AD-associated processes, including dysregulation of the TNF pathway in astrocytes, dysregulation of PD-L1 expression across all cell-types, cellular transport, and lipid dyshomeostasis. Overall, our study provides a novel method to analyze and compare the RNA velocity between AD and non-AD cohorts, and identifies a new set of differentially-spliced AD genes, when previous analyses focused primarily on gene expression changes.

S79. Splicing together the story

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 521. Introme and SpliceVarDB: Predicting and validating splice-altering variation to improve diagnostic yield

Authors:

P. Sullivan[1], V. Gayevskiy[2], R. L. Davis[3], S. Beecroft[4], N. Laing[5], E. Oates[6], A. O'Donnell-Luria[7], M. Pinese[1], M. J. Cowley[1]; [1]Children's Cancer Inst., Sydney, Australia, [2]Walter and Eliza Hall Inst., Melbourne, Australia, [3]Kolling Inst., Sydney, Australia, [4]Univ. of Western Australia, Nedlands, Australia, [5]Harry Perkins Inst. of Med. Res., Nedlands, Australia, [6]Univ. of New South Wales, Sydney, Australia, [7]Boston Children's Hosp. / Broad Inst. / HMS, Cambridge, MA

Abstract:

Variants that affect pre-mRNA splicing can have a substantial impact on the resulting protein. Predicting the impact of coding and noncoding variants on splicing is challenging, particularly in non-canonical splice sites, leading to missed diagnoses in patients. An estimated 9-30% of disease-causing variants affect splicing; however, the number of identified pathogenic splice-altering variants lags behind these estimations. To improve the identification rate of disease-causing splice-altering variants, we have developed SpliceVarDB, a collection of experimentally confirmed splice-altering variants, and Introme, a splicing prediction tool.

Many tools exist to determine if a variant alters splicing; however, the majority are not able to accurately predict the resulting impact on the transcript, which is necessary to determine a splice-altering variant's pathogenicity. To assist in the interpretation of these variants, we have constructed SpliceVarDB, a database of over 21,000 experimentally confirmed splice-altering variants. Additionally, we have curated a set of over 1,000 variants with reported associated splice-altered transcripts and further analysed these to create a set of guidelines to predict the splicing effects of unvalidated splicing variants.

Utilising SpliceVarDB, we built Introme, a splicing prediction tool trained on these experimentally validated splicing variants. Introme uses machine learning to integrate predictions from several splice detection tools, additional splicing rules, and gene architecture features to comprehensively evaluate the likelihood of a variant impacting splicing. We have extensively benchmarked Introme across the SpliceVarDB variants and Introme has outperformed all current leading tools (auPRC: 0.98) for the detection of clinically significant splice variants. Introme has been applied to large cohorts of rare disease and paediatric cancer patients, resulting in numerous genetic diagnoses.

Introme and SpliceVarDB can assist in the improvement of diagnostic yield through a multipronged approach that involves the annotation of splice-altering potential and a resource of validated splicing variants. We intend for SpliceVarDB to become a hub for splicing variant information, allowing users to submit their own validated splicing variants. With the uptake and growth of these resources, we aim to reduce the prediction and validation burden for splice-altering variants.

S79. Splicing together the story

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 522. High-throughput assessment of isoform-specific function by targeting exon-exon junctions in a CRISPR/Cas13 forward screen

Authors:

K. Isaev[1,2], M. Schertzer[2,1], L. Pereira[2], A. Stirn[1,2], A. Das[1,2], H-H. Wessels[2,3], N. E. Sanjana[2,3], D. A. Knowles[2,1]; [1]Columbia Univ., New York, NY, [2]New York Genome Ctr., New York, NY, [3]New York Univ., New York, NY

Abstract:

Alternative splicing is an essential mechanism for diversifying proteins where spliced mRNA isoforms can lead to proteins with distinct functions. A major challenge in characterizing the cellular function of isoforms is the lack of methods to specifically and efficiently downregulate their expression. RNA-targeting RNA-guided (gRNA) type VI CRISPR/Cas13 systems are a recently developed tool to knock down RNA transcripts in mammalian cells with minimal off-target effects, outperforming other methods such as RNAi. While CRISPR/Cas13 has been shown to efficiently target coding regions in mRNA, a transcriptome-wide strategy to knock down distinct isoforms has not been explored. To address this gap, we targeted gRNAs to unique exon-exon junctions in mature mRNA. To assess the ability to target junctions for isoform-specific knockdown, we performed a CRISPR/Cas13 essentiality screen targeting common and unique exon-exon junctions in 844 essential genes in the A375 melanoma cell line. By targeting common junctions, those found in all known isoforms of essential genes, we assessed gRNA efficiency at targeting junctions in a high throughput manner. A linear mixed model was used to assess gRNA dropout in cells across four time points where random effects were estimated for each guide, junction and gene. We found that of the 844 essential genes, guides targeting junctions in 590 (70%) genes were significantly depleted compared to non-targeting controls. Additionally, we identified genes with differential junction effects on proliferation upon knockdown including those in the essential gene *RPS14*. Further, a deep learning model was trained to predict gRNA efficacy. We employed convolutional layers operating on one-hot encoded sequences to learn sequence representations which were flattened and concatenated with a set of non-sequence features and fed into a fully-connected network. Our deep learning model predicted essential gene guide efficacies with a mean pearson correlation of 0.63 suggesting underlying features contain informative signals that influence targetability. Altogether, we establish a strategy and a corresponding computational analysis pipeline to interrogate isoform-specific cellular function in a robust, unbiased and highly expandable manner. Future directions include studying additional functional readouts such as drug resistance.

S79. Splicing together the story

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 523. Investigating X-Linked Dystonia-Parkinsonism signatures in post-mortem brain samples and designing ASOs against XDP signatures in neuronal cell models

Authors:

R. Yadav[1,2,3,4,5], C. A. Vaine[2,5], A. Domingo[1,2,3,4,5], D. Gao[1,2,3,4], S. Shah[2,5], E. B. Penney[2,5], S. Erdin[1,2,3], K. O'Keefe[1], J. Lemanski[1], C. E. De Esch[1], M. McMahon[6], M. Jackson[6], M. Courtney[6], J. Ochaba[6], H. B. Kordasiewicz[6], C. Bennett[6], D. Bragg[2,5], M. Talkowski[1,2,3,4,5]; [1]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [2]Dept. of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [5]The Collaborative Ctr. for X-Linked Dystonia-Parkinsonism, Massachusetts Gen. Hosp., Charlestown, MA, [6]Ionis Pharmaceuticals, Carlsbad, CA

Abstract:

X-linked Dystonia-Parkinsonism (XDP) is an adult-onset neurodegenerative disorder indigenous to the Philippines that exhibits features of both dystonia and parkinsonism. Following de novo assembly of long-read, short-read, and linked-read genome and transcriptome datasets, we previously discovered an expression signature involving aberrant splicing (AS_i32) and intron retention (IR) proximal to an XDP-specific SVA insertion within intron 32 of the *TAF1* gene that was absent in controls. Using patient-specific induced pluripotent stem cell-derived neuronal models, we further demonstrated that this aberrant splicing and reduced *TAF1* expression can be rescued by CRISPR-based excision of the SVA. We have now collected postmortem brain tissues from 21 XDP patients to show that these mis-splicing hallmarks in XDP exist across at least 15 brain regions in cases and are not observed in control brains, with the strongest AS_i32 and IR signatures observed in the prefrontal cortex, cerebellum, and caudate. Motivated by the convergence of in vitro and post-mortem tissue findings, we designed a neural stem-cell based platform to test molecular therapy for this disease using antisense oligonucleotides (ASO) that target XDP-specific sequences in the *TAF1* intron 32, repressing disease-specific splicing events. A total of 81 ASOs were designed to tile the target locus in *TAF1* intron 32, followed by a prioritized and modified set of 37 ASOs. Prioritization was based on their effect on XDP signatures, and the IC50 values. We screened the efficacy of these ASOs by measuring the rescue of IR, suppression of abnormal splicing of AS_i32, and normalization of *TAF1* expression, using global and targeted RNAseq readouts. These analyses led to a series of second-generation enhanced ASOs with improved treatment effects and passed toxicity testing. This study highlights the potential of exploiting patient-specific genome and transcriptome integration as well as targeted therapeutic approaches in seeking precision medicine solutions for a rare and lethal Mendelian disorder.

S80. Therapeutic insights leveraged from preclinical models of disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 526. CRISPR base editing to treat phenylketonuria

Authors:

X. Wang, P. Qu, K. Musunuru; Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Phenylketonuria (PKU), an autosomal recessive disorder caused by mutations in the phenylalanine hydroxylase (PAH) gene, results in the accumulation of blood phenylalanine (Phe) to neurotoxic levels. More than 1,000 PAH variants have been cataloged in PKU patients. Among the five most frequently occurring pathogenic *PAH* variants is the P281L (c.842C>T) variant. Patients homozygous for this variant do not respond to sapropterin (medical therapy), limiting their treatment options. Here our goal is to develop a therapy to correct the P281L variant with a CRISPR adenine base editor (ABE).
PAH is largely expressed in hepatocytes, and ≈10% correction of the pathogenic variant(s) in the liver would in principle be curative in PKU patients. We identified two ABE guide RNAs (gRNAs) with potential to correct the variant via an A>G edit. A challenge in testing for corrective editing of the patient-specific variant by these gRNAs is the lack of readily available human hepatocytes or animal models bearing the variant. We addressed this challenge in two ways.
First, we used prime editing to generate HuH-7 human hepatoma cells homozygous for the P281L variant. In these cells, we screened various ABEs paired with each of the two gRNAs and found that ABE8.8-m in combination with either gRNA showed the highest level of corrective editing (≈50%) with minimal bystander editing. We confirmed on-target activity in primary human hepatocytes into which we introduced the P281L variant sequence via lentiviral genomic integration.
Second, in order to assess for editing activity of ABE/gRNA sets in hepatocytes *in vivo*, we generated an animal model that harbors not only the P281L variant but also the surrounding human sequence context (humanization) that allows for a functional readout of variant correction. We used CRISPR-based homology-directed repair in mouse zygotes to knock in the humanized P281L variant. Homozygous knock-in mice are hypopigmented and have elevated blood Phe levels (>1000 μM/L) compared to wild-type mice (<100 μM/L). We used either (1) dual AAV vectors encoding ABE8.8-m/gRNA or (2) lipid nanoparticles encapsulating ABE8.8-m mRNA and gRNA molecules for delivery into the liver *in vivo*, achieving substantial correction of the P281L variant and phenotypic rescue.
We assessed for off-target editing using two novel methods, ONE-seq and Lenti-seq, which entail screening oligonucleotide libraries—either *in vitro* or incorporated into human hepatocytes via lentiviral genomic integration—for editing by ABE8.8-m/gRNA, followed by targeted amplicon sequencing of candidate sites in hepatocytes, providing critical safety data for an eventual clinical trial application.

S80. Therapeutic insights leveraged from preclinical models of disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 527. 4-Phenylbutyric acid obviates ER stress-induced neurodegeneration in the spinal cord of a mouse model of GM2 gangliosidosis

Authors:

F. Weaver, A. Peek, R. Austin, C. Nurse, S. Igdoura; McMaster Univ., Hamilton, ON, Canada

Abstract:

GM2 gangliosidoses are a group of fatal lysosomal storage diseases, resulting from deficiencies of beta-hexosaminidase A and/or B enzyme activities. Accumulation of GM2 gangliosides in lysosomes within the central nervous system (CNS) causes extensive programmed cell death. Knockout of the hexosaminidase B gene in mice shows progressive neurodegeneration that closely resembles Sandhoff (SD) and Tay Sachs disease (TSD) patients. SD and TSD consist of infantile, juvenile, and late-onset forms. These diseases can present with a multiplicity of symptoms including cognitive and speech impairments, ataxia, and lower motor neuron disease. Late-onset SD and TSD show motor neuron disease in over 40% of patients. We have recently demonstrated the role of ER stress and the unfolded protein response (UPR) during disease pathology in the spinal cord of SD mice. In this study, we have tested the effects of a chemical chaperone 4-Phenylbutyric acid (4-PBA) on the severity and consequences of ER stress in animal models of Sandhoff disease. We administered the drug in drinking water starting at 40 days of age and monitored body weights and behavioural performance until end point. We observed a significant improvement in motor neuromuscular function and life span. Histologically, 4-PBA treatment significantly reduced apoptosis in spinal cord neurons. In comparison to untreated SD mice spinal cords at 80 day of age, we observed a reduced number of cleaved caspase 7-positive neurons and an increased number of cholinergic neurons, with no affect on the severity of astrogliosis. Overall, this study provides strong evidence for the role of chronic ER stress and UPR activation in the spine pathophysiology of LSDs and presents 4-PBA as a potential therapeutic drug for the treatment of Sandhoff disease and other related lysosomal storage disorders.

S80. Therapeutic insights leveraged from preclinical models of disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 528. Potential therapies for Kir7.1 channelopathy

Authors:

M. Kabra[1,2], P. Shahi[1,2], A. Spillane[1], Y. Wang[3,4], D. Sinha[5,2], G. Newby[6,7,8], K. Edwards[2,5], C. Theisen[2,5], S. Saxena[3], K. Saha[2,3], D. Gamm[9,5,2], D. Liu[6,7,8], S. Macdonald[10], S. Hall[10], J. Sabados[10], S. Gong[3,4,11], B. Pattnaik[1,2,9]; [1]Dept. of Pediatrics, Univ. of Wisconsin-Madison, Madison, WI, [2]McPherson Eye Res. Inst., Univ. of Wisconsin-Madison, Madison, WI, [3]Dept. of BioMed. Engineering, Wisconsin Inst. of Discovery, Univ. of Wisconsin-Madison, Madison, WI, [4]Dept. of Materials Sci. and Engineering, Univ. of Wisconsin–Madison, Madison, WI, [5]Waisman Ctr., Univ. of Wisconsin-Madison, Madison, WI, [6]Merkin Inst. of Transformative Technologies in Hlth.care, Broad Inst. of Harvard and MIT, Cambridge, MA, [7]Dept. of Chemistry and Chemical Biology, Harvard Univ., Cambridge, MA, [8]Howard Hughes Med. Inst., Harvard Univ., Cambridge, MA, [9]Dept. of Ophthalmology and Visual Sci., Univ. of Wisconsin-Madison, Madison, WI, [10]Hubble Therapeutics, Boston, MA, [11]Dept. of Chemistry, Univ. of Wisconsin–Madison, Madison, WI

Abstract:

Leber Congenital Amaurosis 16 (LCA16) is a monogenic inherited ocular channelopathy caused by several biallelic point mutations in *KCNJ13*, which affect the retinal pigmented epithelial (RPE) layer of the retina. W53X (c.158G>A; tGg>tAg) mutation in *KCNJ13* results in a truncated Kir7.1 channel. We aim to use two therapeutic approaches to restore the channel function in W53X-LCA16-patient derived iPSC RPE. 1) **Clinical grade AAV5-Kir7.1 HUB101 gene-therapy.** For this, the cells were transduced with AAV5-Kir7.1(10^8, 10^9, and 10^{10}gc/ml) to study the dose-dependent expression of transgene and immune markers expression (qPCR), protein expression (immuno-assays), and function (manual and automated patch-clamp; APC). 2) **Adenosine-CRISPR-base editor** (ABE, ABE8e-spCas9-NG mRNA), delivered by silica nanoparticles (SNP) along with a sgRNA (3:1 molar ratio) to LCA16 fibroblasts and iPSC RPE. On-target/off-target editing efficiency was evaluated by deep sequencing. Electrophysiology was done to evaluate the Kir7.1 channel function in base-edited iPSC RPE. Untreated and isogenic cells were used as reference. The transduced cells showed dose-dependent expression of Kir7.1 transcript. Protein expression confirmed the successful translation of exogenous Kir7.1 and trafficking to the membrane. Electrophysiology showed the rescue of Kir7.1 function. Nanoparticle-mediated delivery of ABE-mRNA in fibroblasts (47%) and post-mitotic hiPS-RPE (20%) established efficient therapeutic base editing. On target Indel mutagenesis (<3%) and deep sequencing of potential off-target sites (<1%) reassured the safety of ABEs. Electrophysiology showed the rescue of channel function in the edited iPSC RPE. **Conclusions•** The expression and functional studies in transduced cells confirmed the clinical use of AAV5 serotype in outer retinal layer. • ABE delivered by SNPs showed the potential to precisely correct the point mutations with reduced or no off-targets, and overcome the adverse effect of AAV-associated immune response. • Functional restoration in transduced and edited cells suggests the potential of this therapies in treatment of pediatric blindness with significant outcomes. • Although, endogenous gene correction via ABE has the upper hand (long-term and permanent change) over exogenous supplementation of Kir7.1 via AAVs, it requires mutation-specific design to correct disease phenotype. In addition, channelopathies like LCA16 require bi-allelic gene correction in the diseased cell to produce a multimeric functional channel. However, AAV-gene augmentation is a universal approach and beneficial for hard to edit mutations due to sequence complexity.

S80. Therapeutic insights leveraged from preclinical models of disease

Location: Conv Ctr/Petree D/West Building

Session Time: Friday, October 28, 2022, 4:00 pm - 5:00 pm

ProgNbr 529. Probucol ameliorates the autoimmune, lipodystrophic and neurodegenerative phenotypes observed in Clec16a KO mice

Authors:

R. Pandey[1], B. Strenkowski[1], T. N. Carneiro[1], M. Bakay[1], H. Hakonarson[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Pediatrics, The Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Introduction: CLEC16A is implicated in multiple autoimmune diseases. We generated Clec16a$^{\Delta UBC}$ mice to address the role of CLEC16A loss of function. Published data suggests a derailed functional link between mitophagy/autophagy/lipophagy, SOCS signaling and CLEC16A, predisposing to the immune dysregulation and the risk of developing an inflammatory phenotype together with neurodegeneration. **Methods:** To address the mechanisms involved, we examined if treatment with drugs specifically targeting mitophagy/SOCS signaling will rescue the defect and prevent the onset of the resulting phenotypes observed in CLEC16A KO mice. We tested the hypothesis that observed perturbations in mitophagy/autophagy, cytokine expression, release and signaling, can be rescued, at least in part by interventions targeting the mitophagy/autophagy, JAK-STAT pathway, including but not limited to SOCS1 and tofacitinib using a drug repurposing therapy approach. **Results:** Turning off Clec16a in adult mice leads to dysregulated mitophagy, severe weight loss, robust autoimmune inflammatory responses, severe neurological symptoms with neuroinflammation and progressive neurodegeneration resembling spinocerebellar ataxia. Treatment with a JAK/STAT inhibitor (tofacitinib) partially rescued the phenotype and improved survival of Clec16a KO mice by modulating ER stress, lipolysis, mitophagy and autophagy. To address the residual phenotype feature, we chose to focus on the ultimate endpoint of mitophagy-clearance of damaged mitochondria to mitigate the negative consequences of mitochondrial damage. In a more tailored approach, we tested the mitophagy enhancer, Probucol (lipid lowering drug) on our whole body inducible Clec16a KO for rescue. Our findings provide evidence in support of targeting mitophagy alone or together with interventions at the JAK-STAT pathway results in phenotype rescue supporting such potential future therapeutic intervention in treating autoimmunity. **Conclusion:** We conclude that in patient populations harboring variants that result in CLEC16A hypofunction, drugs with modulatory effects on mitophagy/SOCS1-JAK-STAT signaling could compensate for the attenuated CLEC16A activity and present formidable candidates for targeted interventions. Thus, our mouse model and results with Probucol and SOCS1-JAK-STAT pathway interventions, serve as a valuable tool for the assessment of therapeutic interventions in patients with autoimmune disease attributed to variants in CLEC16A that impact multiple autoimmune diseases by modulating and rescuing the observed dysregulated mitophagy/autophagy.

S84. A mix of murine models for mechanistic mapping

Location: Conv Ctr/Room 502/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 540. The ENCODE mouse postnatal developmental time course identifies regulatory signatures of cell type maturation

Authors:

E. Rebboah[1], N. Rezaie[1], B. A. Williams[2], J. A. Stamatoyannopoulos[3], H. Ji[4], A. Weimer-Lindeboom[5], M. Snyder[5], B. J. Wold[2], A. Mortazavi[1]; [1]Univ. of California, Irvine, Irvine, CA, [2]California Inst. of Technology, Pasadena, CA, [3]Univ of Washington, Seattle, WA, [4]Johns Hopkins Univ., Baltimore, MD, [5]Stanford Univ., Stanford, CA

Abstract:

Genomic regulation after birth contributes significantly to tissue and organ maturation, but is under-studied relative to existing genomic catalogues of prenatal development in mouse. As part of the final phase of the ENCODE consortium, we generated the first comprehensive bulk and single-cell atlas of postnatal regulatory events across a diverse set of mouse tissues. The collection encompassed seven postnatal time points spanning the human equivalent of childhood through adolescence and adulthood, and focused on adrenal glands, gastrocnemius muscle, heart, hippocampus, and cortex. To allow for allele-specific analyses, we used C57BL6J/Castaneus F1 hybrid mice. Our analysis revealed novel dynamics of cell type composition including identifying new sex-specific cell populations and new commonalities in cell types shared among tissues. We also identify genomic regulatory signatures associated with dynamics of cell type composition, specialisation of sub-cell types, and switching between cell states during postnatal development across 21 different cell types broken down into 68 sub-cell types. We provide an organisational framework to describe TFs that are re-purposed in regulatory signatures of cell type identity in different tissues. Finally, we characterise TFs by their allele-specific expression and transcript isoform choice at the single-cell level. Together, these analyses provide a foundation for understanding the postnatal development of diverse tissues.

S84. A mix of murine models for mechanistic mapping

Location: Conv Ctr/Room 502/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 541. Muscle-specific knock-out of *Sucla2* exhibits differential cellular and *ex vivo* phenotypes as well as striking functional perturbations *in vivo*, providing a novel model of mitochondrial myopathy

Authors:

M. Anderson, P. S. Hafen, A. S. Law, T. B. Meyer, K. D. Fischer, P. J. Gillespie, D. L. McKinzie, J. Brault, B. H. Graham; Indiana Univ Sch. of Med., Indianapolis, IN

Abstract:

Constitutive biallelic pathogenic variants in succinyl-CoA synthetase (SCS) are associated with early-onset mitochondrial encephalomyopathy in human patients. SCS, a heterodimeric TCA-cycle enzyme, catalyzes the interconversion of succinyl-CoA to succinate coupled with substrate-level phosphorylation of either ADP or GDP. This report presents a skeletal muscle-specific conditional knock-out (KO) mouse model of *Sucla2*, the gene encoding the ATP-specific β-isoform of SCS. In this model, the Human Skeletal Actin (*HSA*)-Cre transgene drives recombination of a CRISPR-Cas9 generated floxed allele of *Sucla2* in murine skeletal muscle. Molecular validation confirms over 95% reduction in SUCLA2 protein expression in mutant hindlimb muscles, soleus and extensor digitorum longus (EDL), with subsequent reduction in the α-subunit of SCS, SUCLG1, and elevation of the alternative β-isoform, SUCLG2. Interestingly, despite similar patterns of SCS deficiency, the soleus and EDL exhibit stark differences in other molecular and *ex vivo* phenotypes. For example, while there is no observed difference in mitochondrial DNA (mtDNA) content in the EDL between mutant and wild-type (WT) controls, there is significant mtDNA proliferation in the mutant soleus when compared to WT soleus. Similarly, while there is little to no effect on EDL *ex vivo* muscle function, the soleus, which primarily consists of slow-twitch muscle fibers, exhibits striking perturbations in twitch-force response to a stimulus. In addition to differences in *ex vivo* hindlimb muscle function, the *Sucla2* mutant mice are approximately 50% the size of WT littermates by three weeks of age and remain statistically smaller throughout adulthood. *Sucla2* mutant mice also demonstrate significant changes in *in vivo* muscle function, including widened hindlimb gait, reduced grip strength, and remarkably diminished running wheel activity. Ongoing studies are aimed at investigating the molecular, metabolic, biochemical, and physiological consequences of *Sucla2* deficiency in various muscle types in order to elucidate the mechanisms and roles of the muscle-specific differences as they relate to whole animal phenotypes. Understanding the complexity of these mechanisms will prove critical in the investigation of therapeutic options for *Sucla2*-related mitochondrial encephalomyopathy. Furthermore, this model will provide a useful tool in the investigation into mechanisms of mitochondrial dysfunction, one of the most significant causes of multisystem disease in humans.

S84. A mix of murine models for mechanistic mapping

Location: Conv Ctr/Room 502/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 542. *Psmb5* is essential for mouse early embryo development and possibly for zygote formation

Authors:

Y. Gu; Baylor Coll. of Med., Houston, TX

Abstract:

The Ubiquitin proteasome system (UPS) is a main protein degradation pathway in eukaryotes and defects in this system cause accumulation of excess or damaged proteins, which can lead to abnormal cell function and cell death.
UPS consists of two components: a regulatory component and a catalytic component. The protein encoded by *PSMB5* is one of the catalytic proteins and has chymotrypsin-like activity.
Fragile X-associated Tremor/Ataxia Syndrome (FXTAS) is common in older males with a fragile X CGG premutation allele (55-200 CGG repeats). Major symptoms include intention tremor, cerebellar ataxia and cognitive decline. While the fragile X premutation is relatively common in general population, FXTAS is incompletely penetrant, with many carriers escaping symptoms. Understanding potential genetic modifiers of this phenotype may allow better approaches to therapeutic intervention through mechanistic insights. One candidate modifier, identified through comparison of premutation carrier males with early onset and those without FXTAS using whole exome sequencing and followed by validation studies in a fly model of FXTAS, is *PSMB5* (PNAS 119(22), 2022).
To further confirm *PSMB5's* role in modifying the FXTAS phenotype, we created a *Psmb5* knockout first mouse model using the EUMMR ES cell E09. When *Psmb5* heterozygous males were crossed with *Psmb5* heterozygous females, no homozygote pups were born, demonstrating complete loss of homozygotes during embryogenesis or possibly no development of homozygous zygotes. Among 124 offspring, 94 were heterozygous, 30 were wild type. Curiously, an unexpected increase of heterozygote pups was observed, with a 3:1 Het:WT ratio instead of the expected 2:1 with loss of homozygotes during development. We hypothesize that *Psmb5* KO leads to a very early defect, perhaps with *Psmb5* deficient sperm unable to fertilize *Psmb5* deficient eggs, which instead are selectively fertilized by WT sperm, increasing the number of heterozygotes. We genotyped 52 embryos at ages from E11.5 to E16.5 from *Psmb5* heterozygous males mated with *Psmb5* heterozygous females and found no homozygotes, consistent with our hypothesis.
Analysis of 124 offspring demonstrated weight reduction in both heterozygous males and females in early young adult stages (1-2 months) compared with WT litter mates. Female heterozygotes managed to catch up to wild type animals in body weight after 3 months, but heterozygous males remained leaner than controls at 3 and 4 months. Studies to assess the effects of *Psmb5* reduction on a mouse FXTAS model and creation of conditional *Psmb5* knockout mouse models to investigate tissue specific effects are underway.

S84. A mix of murine models for mechanistic mapping

Location: Conv Ctr/Room 502/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 543. Profiling three-dimensional nuclear telomeric architecture as a biomarker of myelodysplastic syndromes and acute myeloid leukemia in a mouse model

Authors:

M. Gadji[1], R. Boles[2], M. Xu[2], P. Rodrigues[2], A. Pozzo[2], S. Mai[2]; [1]Service of Biological Hematology & Oncology-Hematology (BHOH), Faculty of Med., Pharmacy and Odonto-Stomatology (FMPOS), Natl. Ctr. of Blood Transfusion of Dakar (NCBT/CNTS), Cheikh Anta Diop Univ. of Dakar (UCAD), Dakar Fann BP5002, Senegal, Dakar, Senegal, [2]Dept. of Physiology and Pathophysiology, Rady Faculty of Hlth.Sci., Res. Inst. of Oncology and Hematology (RIOH), CancerCare Manitoba (CCMB), The Genomic Ctr. for Cancer Res. and Diagnosis (GCCRD), The Univ. of Manitoba, Winnipeg, MB, Canada

Abstract:

Introduction: Myelodysplastic Syndromes (MDS) are a group of disorders characterized by cytopenias, with a propensity for evolution into Acute Myeloid Leukemias (AML). However, the cellular and molecular mechanisms causing the transition of MDS to AML remain unknown. In a previous human study, we showed distinct three-dimensional (3D) telomeric profiles specific to patients with MDS and AML that helped subgroup patients based on the severity of telomere dysfunction highlighted in the profiles. These results suggested for the first time a chronological and evolutionary process of telomere dysfunction in MDS leading to transformation to AML (Gadji et al, 2012, Clin Cancer Res; Gadji et al., Adv Cancer Res. 2011). **Method:** In this present study, we investigated the 3D nuclear telomeric profiles based on telomere numbers, telomeric aggregates, telomere signal intensities, nuclear volumes, and nuclear telomere distribution in a unique mouse model that shows progression from MDS to AML "C57BL/6-Tg(Vava1-NUP98/HOXD13)G2Apla/J"2. These hemizygote mice develop MDS with peripheral blood cytopenia and dysplasia and normocellular to hypercellular bone marrow. By 14 months of age, a subset of hemizygotes succumbs to malignant AML or severe anemia and leucopenia. We set up a colony of these mice and the transgenic mice have been bred with C57BL/6NCrL mice. All mice were genotyped and the wild-type littermates are being used as controls. **Results:** We studied ~75 C57BL/6-Tg(Vava1-NUP98/HOXD13)G2Apla/J and ~75 C57BL/6NCrL mice by harvesting samples every month for up to 14 months. The 3D telomere profiles were done comparatively to sex and age matched-control ones. We found a trend of chronological telomere dysfunction in mice that is similar to the human condition. Finally, the two-telomere pathways of progression from MDS to AML already defined with human samples appear to be confirmed by the mouse samples. **Conclusion:** We gain a better understanding of the nuclear processes leading to this transition between MDS and AML that will provide a new basis for new molecular therapeutic strategies aimed at improving the dire prognosis of MDS, AML/MDS and AML. This will, in the future, aid in individualized patient management.

S85. Non-coding variation and cancer

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 546. Non-coding aberrations in mismatch repair genes underlie a substantial part of the missing heritability in Lynch syndrome

Authors:

I. te Paske[1], A. R. Mensenkamp[1], S. van den Heuvel[1], L. Elze[1], K. Neveling[1], S. Baert-Desurmont[2], K. B. M. Claes[3], K. De Leeneer[3], A. Wagner[4], T. J. van Ham[4], M. M. de Jong[5], E. M. Leter[6], R. S. van der Post[1], Y. van Twuijver[1], M. Nielsen[7], N. Hoogerbrugge[1], M. J. L. Ligtenberg[1], R. M. de Voer[1]; [1]Radboud Univ. Med. center, Nijmegen, Netherlands, [2]Rouen Univ. Hosp., Rouen, France, [3]Ghent Univ. Hosp., Ghent, Belgium, [4]Erasmus MC, Rotterdam, Netherlands, [5]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [6]Maastricht UMC+, Maastricht, Netherlands, [7]Leiden Univ. Med. Ctr., Leiden, Netherlands

Abstract:

Background: Individuals with Lynch syndrome (LS) are prone to develop early-onset mismatch repair deficient (dMMR) colorectal- and endometrial cancers due to germline pathogenic variants (PVs) in one of the mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6*, *PMS2*, or deletions affecting the 3' region of *EPCAM*. Current germline diagnostics for LS, after the exclusion of somatic *MLH1*-promoter hypermethylation, include targeted short-read sequencing and multiplex ligation-dependent probe amplification (MLPA) of the coding regions of the MMR genes. In the absence of a germline PV in an MMR gene, the presence of somatic dMMR is investigated. Individuals who remain genetically unresolved after germline and somatic analysis are considered to have Lynch-like syndrome (LLS). For individuals with LLS and their relatives treatment options and surveillance are yet unclear. We applied targeted long-read sequencing (TLR-seq) of the MMR genes to screen for non-coding PVs to explain the observed "missing heritability" in LLS patients.
Methods: We designed long-range targeted PCR amplicons (7-18kb in size) to cover each MMR gene. Germline DNA of LLS individuals (n = 33) was amplified by long-range PCR and pooled prior to single-molecule real-time sequencing on a PacBio Sequel system. Structural variants and rare single nucleotide variants with a predicted effect on mRNA splicing were selected for co-segregation analysis and functional testing in a mini-gene splice assay or a luciferase reporter assay.
Results: TLR-seq identified nine non-coding potential PVs in nine patients. Co-segregation and functional analysis showed that two variants in *MLH1* (c.306+1070C>G, c.306+1001_307-642delinsTA) and three variants in *MSH2* (c.2458+976A>G, c.2459-954A>G, c.212-4_213-3ins366), detected in a total of six individuals (18%), result in aberrant splicing and the introduction of premature stop codons.
One *MLH1* promoter variant (c.-404_-357dup) and a variant in *MSH2* (c.1387-3546_1387-3545ins351; found in *cis* with *MSH2* c.2459-954A>G) remain variants of unknown significance. One *PMS2* variant (c.2276-400G>C) found in two patients did not show altered splicing compared to wildtype.
Conclusion: Our study shows that 18% of individuals with LLS have a germline non-coding PV in an MMR gene that remained undetected by current routine diagnostics. These findings warrant that individuals with LLS, who remain unresolved after germline and somatic analyses, should undergo germline sequencing of the complete loci of (the) MMR genes, preferably by long-read sequencing approaches, to improve LS diagnostics, treatment and cancer surveillance in patients and relatives.

S85. Non-coding variation and cancer

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 547. Single cell chromatin accessibility based deep learning models prioritize functional non-coding genetic variants in colorectal cancer GWAS loci affecting distinct cell types

Authors:

E. Kotler[1], S. Kundu[1], C. Hanson[1], A. Pampari[1], W. Becker[1], J. Huyghe[2], U. Peters[2], M. Snyder[1], W. Greenleaf[1], C. Curtis[1], A. Kundaje[1]; [1]Stanford Univ., Stanford, CA, [2]Fred Hutchinson Cancer Res. Ctr., Seattle, WA

Abstract:

Genome-wide association studies (GWAS) have identified hundreds of risk loci associated with colorectal cancer (CRC), a majority of which reside in non-coding regions with suspected functional consequences on gene regulation. However, deciphering causal variants in these loci remains challenging due to linkage disequilibrium (LD) and the lack of cell-type resolved regulatory landscapes of colorectal tissue. In collaboration with the Human Tumor Atlas Network (HTAN), we recently profiled the chromatin accessibility landscapes and expression profiles of healthy colons, polyps, and CRCs at single cell resolution, providing regulatory maps for over 40 cancer, stromal and immune cell types. Using stratified LD-score regression, we found dozens of significant enrichments for CRC-associated SNPs within regions specifically accessible in individual cell types, implicating them in CRC. To further prioritize likely causal variants that putatively disrupt chromatin accessibility in specific cell types, we developed and trained convolutional neural network models that map regulatory DNA sequence to cell-type resolved, base-resolution scATAC-seq profiles with unprecedented accuracy. Our models account for Tn5-associated bias and allow us to score the functional effects of SNPs based on predicted allelic differences in the magnitude and shape of predicted accessibility profiles. Using a suite of model interpretation frameworks, we highlight predictive sequence motif syntax underlying differential allelic accessibility, enabling the interpretation of transcription-factor binding sites (TFBS) disrupted by predicted high effect size variants. We identified over 50 SNPs in CRC GWAS loci predicted to disrupt accessibility in cell type-specific regulatory elements and examined their likely TFBSs and gene targets. Among these are SNPs predicted to affect epithelial/cancer cell types by activating TFBSs within a super-enhancer residing between the *KLF5* and *KLF12* genes which has been reported to undergo recurrent somatic amplifications in CRC, and near the *GREM1* gene which promotes colonic tumorigenesis. Strikingly, we also found allelic disruptions in non-cancer cell types, such as near the *ZMIZ1* gene in immune cells and *GATA3* in pre-cancer associated fibroblasts (CAFs). Thus, apart from identifying risk alleles, our results suggest a genetic predisposition of both epithelial cells and the tumor microenvironment to the development of CRC. Experimental validation of the effects of these SNPs *in vitro* is underway. Our approach to SNP fine mapping using predictive models of single cell accessibility profiles is generalizable to other diseases.

S85. Non-coding variation and cancer

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 548. Identifying non-coding somatic driver mutations in regulatory elements for pancreatic cancer

Authors:

J. Zhong[1], A. O'Brien[1], J. Hoskins[1], D. Eiser[1], K. Connelly[1], I. Collins[1], T. Shen[2], Y. Zhao[2], L. Wang[3], T. TruongVo[3], A. Wells[4], S. Grant[4], J. Shi[5], L. Amundadottir[1]; [1]Lab. of Translational Genomics, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Rockville, MD, [2]Frederick Natl. Lab. for Cancer Res., Frederick, MD, [3]Lab. of Receptor Biology and Gene Expression, Ctr. for Cancer Res., Natl. Cancer Inst., Bethesda, MD, [4]Children's Hosp. of Philadelphia/Univ. of Pennsylvania, Philadelphia, PA, [5]Biostatistics Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Bethesda, MD

Abstract:

Pancreatic ductal adenocarcinoma (PDAC) is predicted to become the second leading cause of cancer-related deaths in the U.S. by the end of this decade. The identification of somatic driver mutations is fundamental to understanding tumorigenesis and critical for precision oncology. While firmly establishing driver mutations in non-coding, as compared to coding regions of the human genome remains a challenge, Whole Genome Sequence (WGS) datasets from tumor and matched normal samples are enabling progress. In the human genome, open chromatin regions (OCRs) often contain active *cis*-regulatory elements (CREs) bound by transcription factors (TFs) that regulate gene expression of nearby or distant genes. CRE activity can be highly tissue or cell-specific, and enhancers can physically interact with gene promoters at considerable distances, complicating identification of their target genes. To identify somatic non-coding driver mutations in pancreatic cancer, we created genome-scale maps of OCRs, CREs and chromatin interactions to specifically assess enrichment of non-coding somatic mutations (NCSMs) and link them to their respective target genes. We included tumor- and normal-derived pancreatic cancer cell lines, as well as sorted pancreatic ductal and acinar cell populations using assay for transposase accessible chromatin sequencing (ATAC-Seq) and chromatin immunoprecipitation sequencing (ChIP-Seq) (H3K27Ac, H3K27me3, H3K4me1 and H3K4me3). ActiveDriverWGS and DriverPower were then utilized to identify OCRs/CREs enriched with somatic mutations from WGS data from 506 PDAC patients (from ICGC-PACA and PanCuRx). Chromatin interactions between enhancer and promoter regions were assessed by promoter-focused Capture C and H3K27Ac-HiChIP (PANC-1, MIAPaCa-2, sorted acinar and ductal cells) and ENCODE/Roadmap TF binding databases to annotate binding motifs. We identified 314 CREs enriched for somatic mutations (FDR<0.05) that contained 994 NCSMs. Of these, 446 physically interacted with promoters of 338 putative target genes. Moreover, 57 NCSMs disrupted or created new TF motifs. We have assessed gene regulatory effects for the 994 NCSMs using massively parallel reporter assays (MPRA) in three cell lines (PANC-1, MIAPaCa-2 and HEK293T). Of these ~1.5% of mutations significantly altered gene regulatory activity (FDR<0.05) in all three cell lines. Validation and functional assessment of these mutations is currently ongoing. In summary, we have identified potential non-coding driver mutations for PDAC and provide a paradigm for employing a multi-omics approach to decipher complex CREs related to pancreatic cancer biology.

S85. Non-coding variation and cancer

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 549. Oncogenic non-coding RNAs activated by distal enhancers through somatic genome rearrangements in cancer

Authors:

L. Yang; Univ Chicago, Chicago, IL

Abstract:

Somatic genome rearrangements are common in human cancer. The effect of the rearrangements has been studied widely in the context of copy number changes in cancer-related genes and the formation of oncogenic gene fusions. However, the rearrangements can also relocate regulatory elements and the consequences of these events are largely unknown. In recent years, the term “enhancer hijacking” was coined to describe rearrangement events that bring a distal enhancer element close to an oncogene, resulting in its activation. Here, we develop a computational algorithm “HYENA” to identify candidate oncogenes activated by enhancer hijacking based on whole-genome and transcriptome sequencing data. HYENA employs a rank-based regression approach to model gene expression in association with nearby somatic rearrangements, so that it is not sensitive to outliers which is very common in gene expression data. It also computes empirical p values using permutation which serves as a better null distribution. We show that HYENA can detect more known enhancer hijacking genes than other existing algorithms with few false positives. We then systematically analyzed 1148 tumors across 25 tumor types and identified a total of 192 candidate oncogenes activated by distal enhancers. Many candidates are non-coding genes with no known functions. A small nucleolar RNA is activated by recurrent translocations in 15% of thyroid cancers, and three enhancers from the translocation partner chromosome form strong 3D interactions with its promoter. In pancreatic cancer, a long non-coding RNA (lncRNA) is activated by various types of rearrangements in 10% of the tumors. We tested its oncogenic potentials in cell culture models and show that although overexpression of this lncRNA has no effect on cell proliferation in complete cell culture media, it can accelerate cell proliferation in low nutrition media. In addition, overexpression of the lncRNA can promote cell invasion. Since solid tumors often grow in low nutrition condition in vivo, the lncRNA we identified may play important roles in tumorigenesis of pancreatic cancer. Our study highlights the contribution of rearranged regulatory elements such as enhancers in oncogene activation, tumorigenesis and tumor progression.

S86. Novel approaches to increase the diagnostic yield of genetic disorders

Location: Conv Ctr/Room 515/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 552. Using a gene pathogenicity tool 'GenePy' identifies missed biallelic diagnoses in the 100,000 Genomes Project

Authors:

E. Seaby; Univ. of Southampton, Southampton, United Kingdom

Abstract:

The 100,000 Genomes Project (100KGP) diagnosed 25% of recruited patients, but 26% of the diagnoses made were in genes not on the chosen gene panel(s); with many of these being novel *de novo* variants of high impact. However, assessing biallelic variants without the use of a gene panel is extremely challenging due to the sheer number of variants that would require scrutiny. We sought to identify potential missed diagnoses in recessive disease genes independent of the gene panel applied using GenePy - a whole gene pathogenicity scoring approach.
GenePy is a metric that scores all variants called for a given individual, incorporating allele frequency, individual zygosity, and a user-defined deleterious metric (CADDv1.6 applied herein). GenePy then combines variant scores across genes, generating an aggregate gene score for each participant. We calculated GenePy scores for 2862 autosomal recessive disease genes in 68,216 individuals in 100KGP. For each gene, we ranked all participant scores, and scrutinised the data of individuals whose scores ranked amongst the top-5 for each gene, retaining only individuals with a phenotype and without a diagnosis. We assessed these patients' phenotypes for overlap with the phenotype reported for the disease gene for which they were highly ranked. Where phenotypes overlapped, we extracted rare variants in the gene of interested and assessed for phase, ClinVar status, and ACMG classification looking for putative causal biallelic variants.
3191 affected individuals without a molecular diagnosis had a top-5 ranked GenePy gene score and 630/3191 (20%) had phenotypes overlapping with one of the top-ranking genes. In 108/630 (17%) of the phenotype-matched cases, we identified a putative missed diagnosis in a top ranked gene supported by phasing, ClinVar status if available, and ACMG classification. A further 322/630 (51%) of cases have a possible missed diagnosis but may require further functional validation. Scoring GenePy on 2862 recessive genes resulted in curation of 1.2 additional variants per patient with top ranking scores and 0.02 variants for all affected patients in 100KGP.
Applying GenePy at scale has identified potential diagnoses for 430/3191 (13%) of undiagnosed patients who had a top-5 ranked GenePy score in a recessive disease gene, whilst adding minimal additional variants for assessment. We have made available the ranked GenePy scores generated from the 100KGP as a reference dataset which can be applied to individual-level sequencing data for use by the wider scientific community.

S86. Novel approaches to increase the diagnostic yield of genetic disorders

Location: Conv Ctr/Room 515/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 553. Genome reference choice impacts RNA-seq interpretation and rare disease diagnosis

Authors:

P. Goddard, R. Ungar, T. Jensen, K. S. Smith, C. Jin, D. Bonner, S. Marwaha, J. A. Bernstein, M. T. Wheeler, S. B. Montgomery; Stanford Univ., Stanford, CA

Abstract:

Transcriptome analysis is a powerful tool for unraveling molecular mechanisms and improving Mendelian disease diagnostics; however, despite evidence that reference genome selection directly impacts variant identification and diagnostic yield from exome and genome sequencing data, no study to date has systematically assessed the impact of reference genome selection on RNA-seq analysis. We present the impact of the hg19, hg38, and CHM13 reference genomes on expression, splicing, and outlier detection in a rare disease cohort.
We generated paired-end RNA-seq data on blood samples from 330 rare disease patients and family members from the Undiagnosed Diseases Network (UDN), with primarily neurological, musculoskeletal, or immune-related conditions. Data were aligned to the primary genome assemblies, using the GENCODEv35 equivalent gene annotations for hg38.p13, hg19.p13, and CHM13v2. To identify genes with substantial sequence changes between builds, sequence similarity was calculated from pairwise alignment of all exons per gene. We subsequently quantified gene, transcript and splice junction expression, and identified patient-specific expression and splicing outliers.
Pairwise alignment revealed that 99% of genes annotated in both hg19 and hg38 had above 95% sequence similarity across exons, compared to just 12% of genes annotated in hg38 and CHM13. The top 5% of genes with the greatest discrepancies in expression between hg19 and hg38 include 46 genes previously shown to be enriched for variants discrepant between the builds by Li et al (PMID 34129815), and 93 known OMIM genes, including INF2, which underlies a form of Charcot-Marie-Tooth disease. We then investigated how build-induced differences in quantification impacted our interpretations for diagnosed rare disease cases. Seven out of 343 Mendelian disease genes implicated in solved cases across UDN sites had significantly different expression between hg19 and hg38 in our cohort, demonstrating that build can impact RNA-seq guided diagnosis. We further found that overall only 52% of expression outliers and 73% of splicing outliers were consistent between the two builds. We will present similar analyses investigating the impact of switching from hg38 to CHM13.
RNA-seq is a widespread method, yet the understanding of how reference genome selection impacts both quantification results and Mendelian disease diagnosis has not been investigated. We present RNA-seq data on a large number of rare disease patients and demonstrate that in the context of a maximally consistent gene annotation, reference genome selection impacts quantification and clinical interpretation of transcriptomic data.

S86. Novel approaches to increase the diagnostic yield of genetic disorders

Location: Conv Ctr/Room 515/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 554. Variant effect prediction based on custom long-read transcriptomes improves clinical variant annotation

Authors:

R. Salz[1], N. Saraiva-Agostinho[2], E. Vorsteveld[1], C. van der Made[1], J. Allen[2], S. Hunt[2], A. Hoischen[1], P. Hoen[1]; [1]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [2]EMBL-EBI, Wellcome Genome Campus, Hinxton, United Kingdom

Abstract:

Our knowledge of the transcriptome is still incomplete and may result in a failure to detect disease-causing variants. For example, transcripts only expressed under certain conditions such as host-pathogen interactions, are often lacking from reference transcript annotations such as GENCODE and RefSeq. However, in patients with primary immunodeficiencies it could be valuable to annotate variants in these transcripts. Current variant annotation software uses only precomputed pathogenicity prediction scores based on reference transcripts. We thus developed a pipeline based on Ensembl's Variant Effect Predictor (VEP) to annotate variants with custom transcript annotations for downstream prioritization. The input of the pipeline is a sample-specific/non-reference long-read transcriptome in GTF format, variant file(s) (e.g. derived from the exome sequences of patients without a diagnosis) in VCF format, and a reference genome build. The extended version of VEP includes Polyphen-2 to provide custom variant annotations. The input long-read transcriptome contained 37,434 novel transcripts detected through PacBio IsoSeq on peripheral blood mononuclear cells exposed to various immune stimuli. The re-annotation pipeline was tested on all pathogenic variants and variants of unknown significance present in ClinVar. When reannotating 776,866 ClinVar variants with our pipeline, we identified a total of 1,297 variants that were predicted to have a more severe consequence with our pipeline than that predicted with the reference transcriptome. This includes 90 variants in immunodeficiency-related genes, which is more than would be expected by chance (odds ratio=3.63, $p=3x10^{-23}$). Interestingly, 5 of these variants were predicted by the standard VEP annotation to have low severity molecular consequences, but were reported as pathogenic in the literature, demonstrating the added value of the custom transcriptome annotation for pathogenicity prediction. Genetic variant annotation may benefit from long-read sequencing approaches that discover novel transcripts. This benefit can be reaped without extensive bioinformatic knowledge using this pipeline, available at https://github.com/cmbi/VEP_custom_annotations. Our pipeline outputs crucial information for further prioritization of potentially disease-causing variants, and will become increasingly useful as more long-read RNAseq datasets become available.

S86. Novel approaches to increase the diagnostic yield of genetic disorders

Location: Conv Ctr/Room 515/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 555. Automatically constructed pedigrees accurately identify patients at risk for monogenic and polygenic conditions

Authors:

S. Hebbring[1], B. Delgoffe[1], J. Mayer[1], E. Venner[2], D. Muzny[3], R. Gibbs[3], T. Kitchner[1], A. Marcoe[1], E. Stefanski[1], R. Strenn[1], R. Steiner[4], A. Cisler[1], X. Huang[5], D. Page[6], N. Murali[1], M. Brilliant[1], P. Peissig[1]; [1]Marshfield Clinic, Marshfield, WI, [2]HGSC - Baylor Coll. of Med., Houston, TX, [3]Baylor Coll. Med., Houston, TX, [4]UW Sch. of Med. and Publ. Hlth., Marshfield, WI, [5]Univ. of Wisconsin-Madison, Madison, WI, [6]Duke Univ., Durham, NC

Abstract:

A comprehensive family health history is an effective tool for predicting and managing disease risk, including identifying patients for whom genetic testing is clinically indicated. The collection of a clinical family history is often via patient-reported questionnaires sometimes supplemented by active questioning by health care professionals. The small percentage of patients who see genetics professionals typically have family history obtained by active questioning resulting in manual pedigree construction. Unfortunately, the clinical utility of a family history by either method is greatly diminished in practical application due to significant challenges when collecting, storing, retrieving, and disseminating relevant family data. We recently published that family pedigrees could be generated automatically using basic demographic data in an electronic health record (EHR) (e.g., last name and contact information). We hypothesized that combining electronic pedigrees (e-pedigrees) with clinical histories from an EHR across family members could identify patients at risk for monogenic and polygenic diseases. To test our hypothesis, we recruited 2,474 patients into a prospective study to evaluate risk for monogenic forms of cancer, familial hypercholesterolemia, and heart arrhythmia. We further evaluated approximately 20,000 patients retrospectively, to determine whether e-pedigrees could identify patients at risk for 118 more complex polygenic diseases.

Our results confirmed with high confidence that e-pedigrees successfully identified individuals with genetic risk for monogenic forms of cancer ($P=0.0008$), familial hypercholesterolemia ($P=0.0002$), and to a lesser extent, cardiac arrhythmia (0.067); the vast majority with genetic risk unknown to the patients and health system providers prior to e-pedigree construction. In addition to identifying patients at high risk for these rare diseases, e-pedigree data was strongly associated with risk for a wide spectrum of more common polygenic diseases (n=35 diseases, $P<2.8E-4$). In many examples, e-pedigree data remained associated with disease risk even after accounting for measured genetic effects suggesting e-pedigree data could capture additional risk factors shared in families (e.g., environment).

In conclusion, this study provides proof-of-principle that e-pedigrees can capture a family history for many genetic diseases to identify at-risk patients in real-time using only basic information present in the EHR. Furthermore, e-pedigrees are highly structured, scalable to many diseases, and would be amenable for learning healthcare systems to advance precision medicine.

S87. The many ways to make polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 558. The genetic architecture of complex traits and its relevance to polygenic score performance and divergence

Authors:

J. Lachance, A. M. Harris, N. R. G. Carvalho; Georgia Inst. of Technology, Atlanta, GA

Abstract:

Complex traits differ in their genetic architectures, and these differences can affect polygenic score performance. Here, we analyzed a broad panel of 211 complex traits from the UK Biobank. Weights from polygenic scores and dimension reduction approaches were used to identify clusters of traits with similar genetic architectures. Quantitative measures of trait-specific genetic architectures included SNP heritability and Gini coefficients. This latter metric has previously been used by economists to measure income inequality. When applied to genetic data, Gini coefficients quantify whether traits are due a small number of loci of major effect or due to many loci of small effect. We examined three aspects of polygenic score performance for each trait: the correlation between predicted and actual trait values, a portability index which quantifies how well genetic predictions generalize across populations, and a measure of trait divergence between populations. Categorically similar traits (such as BMI and waist circumference) were found to have similar genetic architectures. Traits with high Gini coefficients (i.e., more Mendelian architectures) include Alzheimer's disease, disorders of iron metabolism, and total bilirubin. Traits with low Gini coefficients (i.e., more polygenic architectures) include cannabis use, education attainment, neuroticism, and time spent watching television. Although multiple exceptions exist, highly heritable traits tend to be easier to predict. By contrast, heritability was largely non-informative with respect to the portability of genetic predictions across populations. In general, polygenic score performance was better for quantitative traits than binary traits, a pattern that was due to statistical power considerations. Traits with large population-level differences in polygenic scores include skin pigmentation and hair color. Although Mendelian traits are expected to diverge more between populations than polygenic traits, we did not observe any general trend between Gini coefficients and population-level differences in polygenic scores. Interestingly, lifestyle and behavioral traits tend to have low heritability, small Gini coefficients, as well as poor predictability and portability. Because of this, our results caution against the application of polygenic scores to traits like general happiness, alcohol intake, and average income, especially when polygenic scores are applied to individuals who have an ancestry that differs from the original source population.

S87. The many ways to make polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 559. Increased prediction accuracy for complex traits by combining phenotypic information from known and inferred relatives with polygenic scores

Authors:

M. Moore, P. M. Visscher, L. Yengo, K. E. Kemper, J. Sidorenko; Inst. for Molecular BioSci., The Univ. of Queensland, St Lucia, QLD, Australia

Abstract:

Predictions of complex traits are affected by genetic and environmental factors. Genetic factors can be split into genetic information shared between and within relatives, as well as population-level genetic information. To date, large scale genomic analyses for prediction have predominantly focused on population-level genetic information via the calculation of polygenic scores (PGS). Phenotypes of relatives capture the full heritability and not just that tagged by SNPs. Relative information is commonly incorporated via family history but generally only uses first-degree relatives. This research evaluates a predictor that incorporates phenotypic information from close and distant relatives, by using best linear unbiased prediction (BLUP). The predictor uses estimated genetic relationships between individuals to predict their phenotypes based on the phenotypes of their inferred relatives. Phenotype and genotype data from the UK Biobank (UKB) were used to investigate the predictor for various traits. To incorporate information from inferred relatives the genetic relationship values between participants (N ~ 450,000) were estimated from SNPs. Predictions were calculated for 12 continuous and 1 binary disease trait. A range of traits were selected to evaluate the predictor across various heritabilities (~0.3 - 0.95), sample sizes (~150,000 - 450,000) and data types. The prediction accuracy (R^2) was quantified across UKB participants individually, using a leave-one-out approach (LOO-BLUP) and in conjunction with a PGS. Combining the predictor with a PGS, increased R^2 by 11% and 43% for height and educational attainment, respectively, when compared to the R^2 of the PGS. The predictor has also been extended to include relatives with known relationships to UKB participants, that is, phenotyped but ungenotyped individuals. The prediction of birth weight included 124,013 offspring phenotypes, reported by participants. The R^2 of birth weight predictions for parents increased by ~200% relative to the R^2 of the PGS. For the disease trait diabetes, the R^2 also increased by ~200% when including the predictor in combination with the PGS. Predictions for diabetes included the disease status of 447,705 UKB participants and that of their 427,958 fathers, 440,240 mothers and 376,554 siblings. Hence, we conducted a LOO-BLUP analysis of 1.7M individuals. As bio-bank style databases gather more participants, the density of relatives increases, leading to more information that can be leveraged by this method. This research shows that utilising information from inferred and known relatives via a BLUP can increase accuracy for genomic predictions of complex traits.

S87. The many ways to make polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 560. Addressing overfitting bias due to sample overlap in polygenic risk scoring

Authors:

S. Jeong[1], M. Shivakumar[2], K. Nho[3], S. Lee[1], D. Kim[2]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Indiana Univ Sch. of Med., Indianapolis, IN

Abstract:

Polygenic risk score (PRS) is an important tool for complex disease research. In constructing PRS, publicly available GWAS summary data from large consortia are commonly used. Since these consortia typically include most of the existing GWAS, the GWAS summary data often have samples of the target dataset that will be used to evaluate the performance and utility of PRS. However, the potential bias due to the sample overlap has not been extensively investigated, and no method has been developed to address the bias. Using coronary artery disease (CAD) in UK Biobank data as a disease (9,323 cases and 368,586 controls), we first conducted extensive evaluations of the potential bias. We prepared train-test set splits with increments of overlap proportions. There was a maximum 4.1% inflation on AUROC of PRS (0.788 to 0.820) as the proportion increased. The result showed that even with as little as 11.8% of overlap in the training sample, PRS could result in significant overfitting. To address the overfitting bias, we developed a novel method, overlap-adjusted PRS (OA-PRS). With the assumption that overlapped samples in the target dataset are known, the proposed approach excludes the overlapped samples from the GWAS summary data by inverting the meta-analysis procedure and uses the overlap excluded GWAS summary data to construct PRS. Simulation studies showed that the proposed overlap exclusion procedure could produce nearly identical effect size estimates (>0.99 correlation) from the individual level data with known overlap samples excluded, and maintain the same PRS AUROC level. UK Biobank CAD analysis also showed that the proposed OA-PRS could address the overfitting bias. By investigating the overfitting bias due to the sample overlap and developing a method to address it, our work will further facilitate the utility of PRS with the availability of multiple biobanks.

S87. The many ways to make polygenic risk scores

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 561. Potential clinical utility of polygenic risk scores in disease prognosis. The glass is half full

Authors:

J. Wanner[1,2], N. Mars[2], Z. Yang[2], FinnGen, INTERVENE, S. Ripatti[2], A. Ganna[2], H. Heyne[1,2,3]; [1]Hasso Plattner Inst., Digital Hlth.Ctr., Potsdam, Germany, [2]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, [3]Hasso Plattner Inst., Mount Sinai Sch. of Med., New York, NY

Abstract:

Most research on polygenic risk scores (PRSs) focuses on incidence of diseases, not their prognosis. In the clinic, prediction of disease outcomes happens however more frequently in at-risk than in healthy individuals. Here, we explore the utility of polygenic risk scores for disease prognosis using electronic health records of 392,649 Finns. First, we investigate clinical settings where a patient presents with an unspecified symptom that may be a sign of a more severe disease. For example, we explore the clinical case of an unspecified first seizure at risk to develop epilepsy and unclear stomach complaints (irritable bowel syndrome, IBS) at risk to develop an autoimmune inflammatory bowel disease (IBD). In 7,770 individuals with IBS, we found that individuals with the top 5% of IBD-PRS showed an increased relative risk (HR 2.3, 95% CI 1.57-3.40, $p=2.19*10^{-5}$) of developing IBD after 10 years compared to individuals with lower IBD-PRS with a larger effect in young adults (age < 40). Furthermore, we show that the cumulative lifetime incidence of IBD reaches ca 22% in individuals with the top 5% of IBD-PRS and a previous IBS diagnosis, while individuals without a previous IBS diagnosis remain at a lower risk (ca 10%). By an increase in absolute risk after a first milder or unspecific event we thus highlight ways to potentially improve the clinical utility of disease incidence PRS in stratifying risk of severe disease. Additionally, we explored the potential of PRSs of disease incidence to predict prognosis such as disease severity. In the example of IBD, we used IBD-related surgery as a proxy for disease severity. Here the top 5% of IBD-PRS show an increased risk of undergoing total colectomy compared to the rest of the cohort (HR 1.52, 95% CI 1.15-2.02, $p=3.78*10^{-3}$), indicating similarities in the genetic architecture for incidence and progression of IBD. We investigate the predictive value of incidence PRSs on disease prognosis in multiple other diseases such as atrial fibrillation, epilepsy and prostate cancer. We find a substantial correlation between genetic factors influencing disease incidence and disease severity for some but not all diseases and in some cases attenuation of PRS effects due to selection bias. Analyses are replicated in a biobank network (n=1.2 Million). We thus present potential applications of PRSs for predicting factors of disease prognosis.
Grants:Horizon 2020(101016775)

S88. Unwrapping the role of chromatin in neurodevelopmental disorders

Location: Conv Ctr/Petree C/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 564. A framework for summarizing chromatin state annotations within and identifying differential annotations across groups of samples

Authors:

H. Vu, Z. Koch, P. Fiziev, J. Ernst; Univ. of California, Los Angeles, Los Angeles, CA

Abstract:

Genome-wide maps of epigenetic modifications are powerful resources for non-coding genome annotation. Maps of multiple epigenetics marks have been integrated into sample-specific chromatin state annotations for many samples of various cell or tissue types. With the increasing availability of multiple chromatin state maps for biologically similar samples, there is a need for methods that can effectively (1) summarize the information about chromatin state annotations within groups of samples and (2) identify chromatin state differences across groups of samples at a high resolution.
To address this, we developed CSREP, which takes as input chromatin state annotations for a group of samples and then probabilistically estimates state assignment probabilities at each genomic position, which it then summarizes with the maximum-probability chromatin state at each position for the group. CSREP uses an ensemble of multi-class logistic regression classifiers to predict the chromatin state assignment of each sample given the state maps from all other samples with shared biological properties. The difference of CSREP's summary probability assignments for two groups can be used to identify genomic locations with differential chromatin state patterns.
Using groups of chromatin state maps of a diverse set of cell and tissue types, we demonstrate the advantages of using CSREP to summarize chromatin state maps. We show that CSREP is better able to predict the chromatin state of held out samples in a group than the counting-based baseline approach. We also show that CSREP can identify biologically relevant differences between groups at a high resolution, such as predicting cell-type-specific peaks of histone modification marks. The CSREP source code, and the summary chromatin state maps for 11 sample groups from Roadmap Epigenomics project, and 75 sample groups from Epimap is available at http://github.com/ernstlab/csrep.

S88. Unwrapping the role of chromatin in neurodevelopmental disorders

Location: Conv Ctr/Petree C/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 565. *De novo* mutations in replication-independent histone genes and an unexplored class of rare pediatric Mendelian syndromes

Authors:

E. Lubin[1], L. Bryant[2], J. Aicher[3], D. Li[2], E. Bhoj[2]; [1]Perelman Sch. of Med./Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Univ. of Pennsylvania, Philadelphia, PA

Abstract:

Mutated replication-independent (RI) histone genes are an emerging cause of rare pediatric Mendelian syndromes. Recurrent “hot spot” somatic mutations in members of the histone H3.3 family were first identified in large cohorts of patients with pediatric gliomas. Germline mutations in multiple histone genes encoding distinct proteins have subsequently been found to cause neurodevelopmental and neurodegenerative disorders that may act through a targetable dominant-negative mechanism. These RI histone mutations eluded detection by exome and genome sequencing methods, despite patients’ profound phenotypes, because of the high sequence homology between histone genes. Thus, we employed an innovative approach using population-level data, specifically gnomAD constraint metrics and GTEx expression data, to predict that five previously unreported histone-encoding genes (*H2AFV/H2AZ2, H2AFY/MACROH2A1, H2AFY2/MACROH2A2, H2AFZ/H2AZ2* and *H1F0/H1-0*) are most intolerant to mutations, rendering them putative disease candidate genes. Working with our international network of collaborators and GeneMatcher, we have subsequently validated our approach by identifying cohorts of patients with mutations in three of those five disease candidate genes (*H2AFY/MACROH2A1, H2AFY2/MACROH2A2, H2AFV/H2AZ2*). These patients share overlapping neurodevelopmental phenotypic traits, including intellectual disability and developmental delay. To delineate the pathogenesis of these mutations in a neurodevelopmental context, we are developing a pipeline in which we generate isogenic iPSC lines harboring the patients’ mutations and then quantify the resulting epigenetic dysregulation. Combining our innovative candidate gene prediction project with our patient-driven approach to interrogating disease-relevant mutations allows us to both overcome the limitations of traditional sequencing methods and build evidence for a potential unifying model by which de novo heterozygous missense mutations in RI histone genes act through a shared, therapeutically targetable dominant-negative mechanism.

S88. Unwrapping the role of chromatin in neurodevelopmental disorders

Location: Conv Ctr/Petree C/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 566. Epigenetic mechanisms of growth retardation in Kabuki syndrome 2 and correction of disrupted H3K27 methylation as an approach to treatment.

Authors:

C. Gao[1], W-Y. Lin[1], L. Boukas[1,2], K. D. Hansen[2], R. Riddle[1], H. T. Bjornsson[3,1], J. A. Fahrner[1]; [1]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Johns Hopkins Univ. Sch. of Publ. Hlth., Baltimore, MD, [3]Univ. of Iceland, Reykjavik, Iceland

Abstract:

Kabuki syndrome 2 (KS2) is a Mendelian disorder of the epigenetic machinery (MDEM) resulting from germline hemi- or heterozygous pathogenic variants in *KDM6A*. KDM6A is a demethylase (eraser) of H3K27 di- and tri-methylation. We hypothesize that perturbations to the H3K27 epigenome lead to altered gene expression in multiple tissues, which manifests clinically as the characteristic features of KS2: neurological dysfunction, craniofacial dysmorphisms, and growth retardation. Notably, the phenotype of KS2 closely resembles Kabuki syndrome 1 (KS1); yet, the causative gene in KS1, *KMT2D*, encodes a methyltransferase (writer) targeting H3K4. Previously, we implicated precocious differentiation of chondrocytes as a mechanism of growth retardation in KS1. In $Kmt2d^{-/-}$ chondrogenic cell lines, H3K4me3 levels were decreased at the promoter of *Shox2*, critical to mouse chondrocyte development and long bone growth. RNA-seq further revealed ~1000 genes to be differentially expressed compared to genotype controls. Here we focus on KS2. Specifically, we characterize an *in vivo* mouse model that recapitulates the growth retardation phenotype, and we use an *in vitro* chondrogenic model to investigate the underlying molecular mechanisms. We demonstrate that heterozygous $Kdm6a^{tm1d/+}$ mice, generated using the knockout-first approach, have decreased body length and weight, shortened femurs and tibias, and decreased bone area by high-resolution micro-CT. Growth plate abnormalities are apparent histologically. We created $Kdm6a^{-/-}$ chondrogenic cell lines by CRISPR/Cas9 genome editing. In comparison to $Kdm6a^{+/+}$ controls, $Kdm6a^{-/-}$ lines show enhanced chondrocyte differentiation by Alcian blue staining and significantly increased expression of the chondrogenesis marker *Col10a1*. Ongoing RNA-seq and H3K27me3 ChIP-seq studies will reveal pathways involved in growth retardation in KS2 and potentially other MDEMs. We are particularly intrigued by the striking phenotypic resemblance between KS1 and KS2. We hypothesize that this may result from similar disruptions to the chromatin state at shared effector genes, albeit through different epigenetic mechanisms involving H3K4me3 and H3K27me3, respectively. Finally, we are investigating the potential of epigenetic modulating agents such as Ezh2 inhibitors, which block the writer of H3K27me3, to reverse molecular and phenotypic effects due to loss of Kdm6a function both *in vitro* and *in vivo*. Ultimately, these studies will further our understanding of growth abnormalities in KS2, and bear promise for treating the multisystemic sequelae of KS2 and other MDEMs by addressing the molecular etiology of disease.

S88. Unwrapping the role of chromatin in neurodevelopmental disorders

Location: Conv Ctr/Petree C/West Building

Session Time: Saturday, October 29, 2022, 9:00 am - 10:00 am

ProgNbr 567. Uncovering the role of EZH1 in neural development and neurodevelopmental disorders

Authors:

C. Gracia Díaz[1,2], Y. Zhou[2], Q. Yang[1], S. Zhang[1], G. Otrimski[2], E. Heller[1], H. Song[1], M. Guo-Li[1], N. Akizu[2,1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Abstract:

The disruption of chromatin modifiers during cortical neurogenesis is one of the underlying causes of neurodevelopmental and intellectual disability disorders. We recently found *EZH1* loss and gain of function (LOF and GOF) variants as the genetic basis of overlapping neurodevelopmental disorders. Enhancer of Zeste Homologue 1 (EZH1), is one of the two histone H3 lysine 27 (H3K27) mammalian methyltransferases, and part of the Polycomb Repressive Complex 2 (PRC2). Its paralogue, EZH2, has an essential role in maintaining transcriptional repression of non-lineage specific genes during development. In a neurodevelopmental context, EZH2 is highly expressed in dividing cells, and its dysfunction leads to defects in neural progenitor proliferation and fate specification. In contrast, the contribution of EZH1 to transcriptional silencing and neural development or disease is poorly understood. By interrogating publicly available transcriptomic datasets we found that EZH1 is expressed in both the developing and adult nervous system, highlighting the likely relevance of EZH1 in the developing and adult human brain. To test this hypothesis, we generated neurodevelopmental models derived from EZH1 LOF and GOF human pluripotent stem cells (hPSC). Our monolayer neuronal differentiations showed that EZH1 LOF-derived neurons are less mature than wild-type neurons. Consistently, using cortical brain organoids, we uncovered that EZH1 is necessary for the coordination of cortical neurogenesis timing. Specifically, EZH1 LOF cortical organoids showed delayed neurogenesis of lower- and upper-layer projection neurons. In contrast, hyperactive GOF EZH1 neurons exhibited increased mature neuronal features, such as longer neurites. Furthermore, cortical organoids revealed that EZH1 GOF leads to premature neurogenesis specifically affecting upper-layer cortical neurons. Our current work is focused on deciphering the function of EZH1 and the effects of pathogenic variants in the epigenetic transcriptional regulation of the nervous system. Our work uncovers an essential role of EZH1 in cortical neurogenesis timing, and how when its function is dysregulated, it results in neurogenesis timing asynchronies and overlapping neurodevelopmental disorders.

S89. Detection and effect of CNVs in general populations

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 570. A high-resolution map of copy number and structural variation in Qatari genomes and their contribution to quantitative traits and disease

Authors:

E. Aliyev[1], A. Visconti[2], N. Syed[1], T. Aliyev[3], R. Razali[1], M. Ghorbani[1], H. Naeem[1], A. Belkadi[4], G. Thareja[1], N. Rossi[2], W. Aamer[1], K. Suhre[4], Y. Mokrab[1], M. Falchi[5], K. Fakhro[6]; [1]Sidra Med., Doha, Qatar, [2]Dept. of Twin Res. & Genetics Epidemiology, King's Coll. London, London, United Kingdom, [3]CERN, Meyrin, Switzerland, [4]Weill Cornell Med., Doha, Qatar, [5]King s Coll. London, London, United Kingdom, [6]Sidra Med., Doha, Qatar, Qatar

Abstract:

In this study, we developed an ensemble pipeline for the calling structural variants (SVs) from WGS data in about 6000 samples and generated a reference atlas of SVs that cataloged almost 100,000 SVs in the Qatari population, which helped to elucidate population structure at high resolution and provide a genetic explanation to the missing heritability in some diseases observed in this population, critical to the future of personalized care in Qatar. During this study, we developed a comprehensive, home-grown pipeline that can detect deletions, duplications, inversions, translocations, and insertions from WGS data across large cohorts (e.g., 6,000 individuals), and developed tools to annotate these SVs to estimate the biological impact of each variant. Our analyses revealed diverse mutational patterns among SVs, and strong selection acting against reciprocal dosage changes. Prepared a "map" of SVs in the Qatari population, which helped to provide a genetic explanation to some common and rare diseases observed in this population and will help the medical community provide personalized care in Qatar. Some of our findings can be integrated in clinical practice, for diagnostic/genetic counselling purposes. This not only has implications for human health and disease but also represents the backbone structure for a Qatari-specific reference genome. Results of this study will be widely shared with the research community, for use in global cohorts, therefore increasing our understanding of the impact of SVs on population health and disease. Overall, this study played a significant role in building home-grown capacity for genome structure analysis in Qatar - a very challenging yet rewarding discipline - which forms a strong nucleus upon which the country will continue to build, as it charts a future of excellence in delivering Precision Medicine to its growing population.

S89. Detection and effect of CNVs in general populations

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 571. Copy number variants differ in prevalence across ancestral populations

Authors:

L. M. Schultz[1], A. Knighton[2], G. Huguet[3], M. Jean-Louis[3], Z. Saci[3], J. Mollon[4], E. Knowles[4], D. C. Glahn[4], S. Jacquemont[3], L. Almasy[2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]CHU Sainte-Justine, Montreal, QC, Canada, [4]Boston Children's Hosp., Boston, MA

Abstract:

Rare copy number variants (CNVs) have been implicated in a variety of disorders. Given that frequencies of SNPs and other genetic variants differ across ancestral populations, we investigated whether there are also differences in CNV prevalence. We used PennCNV and QuantiSNP to call CNVs (≥ 50 kB pairs) for white British (WB; *n* = 385,636), other European (EUR; *n* = 51,334), South Asian (SAS; *n* = 8,848), and African (AFR; *n* = 8,447) individuals from the UK Biobank whose genetic ancestry we inferred using KING.

Differences between expected and observed counts of CNV carriers were analyzed via chi-square tests for independence followed by post-hoc comparisons of standardized residuals; all *p*-values were FDR-corrected.

Examining the prevalence of individuals possessing at least one deletion (DEL) or at least one duplication (DUP), there were more SAS DEL carriers ($p < .0000001$) and fewer AFR DUP carriers ($p < .0000001$) than expected. When we limited our analyses to CNVs under constraint (i.e., ≥ 2 intolerant genes based on LOEUF), we found fewer AFR DEL ($p = .0004$) and DUP ($p < .000001$) carriers and fewer SAS DEL ($p = .0002$) and DUP ($p = .001$) carriers than expected. Focusing on recurrent DELs ($n = 50$) and DUPs ($n = 60$), there were fewer AFR DEL ($p = .0002$) and DUP ($p = .000001$) carriers than expected. Further limiting the focus to constrained recurrent CNVs, there were fewer SAS ($p = .0002$) and AFR ($p = .0003$) DEL carriers and fewer AFR DUP ($p = .00007$) carriers than expected. These results may reflect ancestral differences in flanking tandem repeat sequences as well as Eurocentric biases in defining recurrent CNVs and deriving constraint measures.

To assess whether individual recurrent CNVs display ancestral differences in prevalence, we identified 7 DELs and 9 DUPs that each occurred at least 275 times in the pooled dataset or at least 5 times in the SAS or AFR cohort. Among these CNVs, we observed differences relative to expectation for 15q11.2 DEL (WB+, SAS-, AFR-; *p*'s < .007) and DUP (SAS+; $p = .006$), 2q13 *NPHP1* DEL (AFR-; $p = .003$) and DUP (WB-, EUR+, SAS+, AFR+; *p*'s < .03), *ZNF92* DEL (WB-, SAS+; *p*'s < .000003), and 15q13.3 BP4.5-BP5 *CHRNA7* DUP (WB+, SAS-, AFR-; *p*'s < .05).

We observed up to two-fold differences in the rates of individual CNVs across ancestry groups. Several of the recurrent CNVs that differ across populations have been linked to neurodevelopmental and medical outcomes in studies limited to white British individuals. Given the differences in CNV prevalence across populations and the potential for Eurocentric bias, it is imperative that future CNV studies include AFR and other non-EUR ancestral populations.

S89. Detection and effect of CNVs in general populations

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 572. The impact of rare coding CNVs in 197,306 UK Biobank exomes

Authors:

J. Fu[1], A. Smirnov[2], R. Collins[1], S. Lee[2], M. Walker[2], I. Wong[1], D. Benjamin[2], A. Sanchis-Juan[2], K. Karczewski[2], H. Brand[1], E. Banks[2], L. Gauthier[2], M. Babadi[2], M. Talkowski[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA

Abstract:

Exome sequencing (ES) has been generated for millions of individuals across international biobanks, human disease studies, and clinical diagnostic screens. The vast majority of ES-based studies have been restricted to short variants (single nucleotide variants and small indels), which are routinely detected using established methods. Copy number variants (CNVs) are important contributors to a broad spectrum of human diseases, but the relative lack of standardized CNV detection tools for ES data has presented a major barrier to their routine assessment in large-scale ES-based studies. To overcome this challenge, we developed GATK-gCNV, a tool for calling CNVs from read depth data using Bayesian inference. We extensively benchmarked GATK-gCNV using over 7,000 samples with matched genome and ES data, demonstrating an ability to capture 86% of CNVs detectable by gold-standard genome-based approaches while maintaining a false positive rate <10%. We applied GATK-gCNV to discover rare (<1% site frequency) coding CNVs in ES data from 197,306 UK Biobank (UKBB) participants. This atlas comprised 38,731 rare coding autosomal CNVs larger than two exons, with 64% of individuals carrying at least one such variant. We further found very strong correlations between CNV rates per gene and orthogonal measures of genic constraint (*e.g.*, Pearson's correlation r=0.97 for deletions vs loss-of-function observed/expected upper bound fraction [LOEUF], r=-0.94 for duplications vs probability of triplosensitivity [pTS]). Initial phenome-wide association analyses, which included assessment of 79 genomic disorder loci, recapitulated known associations with phenotypic traits (*e.g.*, body mass index with 16p11.2 deletions) and further identified focal CNV associations that included deletions of *PDZK1* with urate levels, deletions of *HBA1, HBA2, HBM, HBQ1* with blood traits, and duplication of *CTS3* with decreased estimated glomerular filtration rates. In conclusion, GATK-gCNV demonstrates sufficient sensitivity and specificity to discover rare CNVs in routine research and clinical applications of ES data. We further provide here a resource of rare coding CNVs in the UKBB for use by the research community with intriguing rare CNV trait associations that are not fully captured by routine analysis of short variants.

S89. Detection and effect of CNVs in general populations

Location: Conv Ctr/Concourse Hall F/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 573. Analysis of gene dosage shows that the majority of the coding genome impacts

Authors:

G. Huguet[1], E. Douard[1], T. Renne[2], C. Proulx[1], C. Poulain[2], J. Mollon[3], L. Schultz[4], M. Jean-Louis[1], Z. Saci[1], Z. Pausova[5], T. Paus[1], L. Almasy[4], D. Glahn[3], S. Jacquemont[1]; [1]CHU Sainte-Justine, Montréal, QC, Canada, [2]Montreal Univ., Montréal, QC, Canada, [3]Boston Children's Hosp., Boston, MA, [4]Children's Hosp. of Philadelphia, Philadelphia, PA, [5]The Hosp. for Sick Children, Toronto, ON, Canada

Abstract:

Background: Pathogenic copy number variants (CNV) are identified in 10-15% of patients referred for neurodevelopmental disorders. We have previously shown that models using intolerance to loss of function scores (LOEUF) can estimate the effect size of any CNV on intelligence quotient (IQ) with an accuracy close to 80%. However, these models assumed that the effects of CNV were mainly driven by negative effects of intolerant genes on cognitive ability. **Knowledge gap:** The proportion of the coding genome influencing cognitive ability when affected by gene dosage remains unknown. **Hypothesis:** By partitioning the genome based on intolerance scores, we can infer effect sizes of individual genes on cognitive ability.
The aim: Provide estimates of effect size on cognitive ability of all coding genes based on their constraint score values.
Methods: We identified and filtered all CNV >50kb from microarray data using PennCNV and QuantiSNP, in 258,292 individuals from unselected populations and 5,104 individuals from autism cohorts. Cognitive ability was measured using different assessments. We partitioned the genome in 38 overlapping categories of genes with increasing levels of intolerance to loss of function measured by LOEUF. We then performed 38 linear models to estimate the mean effect size of a gene within each category.
Results: Out of the 18,451 genes with a LOEUF value, 38.2% and 67.3% were fully deleted or duplicated, respectively. We computed estimates for 38 categories with two methodologies, meta-analyses and pooled cohorts. The observations by both methods were similar. Genes showed significant effects on cognition in 17 and 19 out of the 38 categories when deleted or duplicated, respectively, suggesting that half of the genome modulates cognition. For deletions, effect sizes were largest for intolerant genes and decreased progressively from LOEUF<1,showing significant negative effects ranging from -0.75 to -3 IQ points. The same pattern with smaller effect sizes was observed for duplications. We also identified 2 groups of genes (LOEUF≥1) increasing cognition when deleted and 5 groups when duplicated. Sensitivity analysis showed that these findings were robust. We provide a detailed map of brain cell type and the developmental expression profile for each significant category.
Conclusion: Results suggest that gene dosage influences cognition across most of the coding genome. While any gene with even the slightest intolerance may negatively affect cognition, another group of tolerant genes positively affects this trait. We will add the models in the predictive tool (https://cnvprediction.urca.ca/) to improve the interpretation of CNV in the clinic.

S90. Expanding roles of repeats

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 576. Polymorphic short tandem repeats make widespread contributions to blood and serum traits.

Authors:

J. Margoliash[1], Y. Li[1], A. Massarat[1], S. Fuchs[2], A. Goren[1], M. Gymrek[1]; [1]Univ. of California, San Diego, CA, [2]Inst. of Endocrinology, Sheba Med. Ctr., Tel Hashomer, Ramat Gan, Israel

Abstract:

Short tandem repeats (STRs), genomic regions each consisting of a sequence of 1-6 base pairs repeated in succession, represent one of the largest sources of human genetic variation. Polymorphism in the number of repeats at thousands of STR loci has been associated with gene regulation. However, many STR effects are not captured well by standard genome-wide association studies (GWAS) or downstream analyses that are mostly based on single nucleotide polymorphisms (SNPs).To study the involvement of STRs in genetic traits, we imputed genotypes for 445,735 autosomal STRs into SNP data from 408,153 UK Biobank participants and tested for association with 44 blood and serum biomarker phenotypes. We used two fine-mapping methods, SuSiE and FINEMAP, to identify GWAS signals putatively caused by STRs and estimate that STRs drive 4.3-9.7% of GWAS signals for these traits. We additionally evaluated the concordance and robustness of these two fine-mapping methods. We find that they have an agreement rate of 32% amongst potentially causal STRs, and that only 53% of STRs initially fine-mapped by both methods remain candidate causal variants under a range of alternative fine-mapping conditions, highlighting that current fine-mapping is highly sensitive to parameter and algorithmic choices. Overall, we identify 118 high-confidence STR-trait associations that pass all of our fine-mapping tests. We performed replication analyses in the Black, South Asian and Chinese ancestry groups in the UK Biobank and find that confidently fine-mapped STR associations show concordant directions of effects in non-White groups more often than STR associations which did not pass fine-mapping (per population p-values ≤ 0.002). Our high confidence STR-trait associations implicate STRs in some of the strongest hits for multiple phenotypes. For example, we identify a known trinucleotide STR in *APOB* as causally mediating LDL cholesterol and apolipoprotein B concentrations (association p-values = 2e-235, 1e-279). As an additional example, our candidate causal variants include a novel association between a CGG repeat in the promoter of *CBL* and the traits platelet count and platelet crit (respective p-values 4e-83, 6e-103), which we corroborate with orthogonal evidence from the Geuvadis and GTEx cohorts showing that variation in the length of this STR affects *CBL* expression (p=3e-5 after multiple hypothesis correction). Together, our results suggest polymorphic tandem repeats make widespread contributions to complex traits, we provide a set of stringently selected candidate causal STRs, and we demonstrate the need to routinely consider a more complete view of human genetic variation in GWAS.

S90. Expanding roles of repeats

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 577. Phenotypic effects of common coding copy-number variation

Authors:

M. Hujoel[1,2], M. Sherman[3,1,2], A. Barton[4,1,2], R. Mukamel[1,2], P-R. Loh[1,2]; [1]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Massachusetts Inst. of Technology, Marblehead, MA, [4]Harvard Univ., Cambridge, MA

Abstract:

The effects of structural variants (SVs) on human phenotypes are understudied due to the difficulty of genotyping SVs in large biobank cohorts. We developed a new haplotype-informed method to detect and genotype common copy-altering SVs from whole-exome sequencing (WES) read-depth data available for 454,799 UK Biobank participants. This approach first leverages familial relationships to discover genomic regions with heritable read depth (suggesting the presence of inherited copy-number variation) and then leverages haplotype-sharing among the full cohort to refine copy number estimates derived from read depth for all samples.
Applying this approach to search for copy-altering SVs affecting a broad set of genomic regions (including 100bp bins, exons, and previously reported CNVs), we detected 41,966 such regions with heritable WES read depth and consistent WGS read depth available for a subset of samples. Of 18,765 annotated genes, 2,480 (13%) had evidence of an SV overlapping a coding exon. As expected, these genes tended to have lower predicted probability of loss of function intolerance (average pLI=0.17 across genes impacted by an SV versus 0.24 across all genes). Association and fine-mapping analyses of these SVs with 56 quantitative traits identified 337 independent associations ($P < 5 \times 10^{-8}$) that could not be explained by linkage disequilibrium with any nearby SNP. These associations involved not only simple deletions and duplications but also multi-copy polymorphisms and tandem repeat variants, which read-depth analysis could also detect. A 584bp deletion spanning exon 6 of *CTRB2* (allele frequency 7.6%) associated with lower HbA1c levels ($p=8.0 \times 10^{-19}$) and decreased risk of type 2 diabetes (OR=0.84 (0.81-0.88), $p=1.6 \times 10^{-17}$). A 39bp coding repeat within *GP1BA*, which encodes platelet glycoprotein Ib alpha chain, associated with increased platelet counts ($p=1.9 \times 10^{-84}$) and other platelet phenotypes. An intronic trinucleotide repeat expansion within *TCF4* (transcription factor 4) associated with altered levels of multiple serum biomarkers (creatinine, urate, HbA1c, and testosterone levels), blood counts, and BMI and FEV1/FVC ratio ($p=2.4 \times 10^{-16}$ - 6.9×10^{-10}). This repeat expansion has previously been implicated in Fuchs corneal dystrophy; here we further observed an association with increased risk of cataracts ($p=3.5 \times 10^{-9}$; OR for top percentile=1.31 (1.19-1.44)). These results indicate the potential of haplotype-informed techniques to enable further insights into the phenotypic impact of structural variation by more powerfully leveraging short-read sequencing data sets generated to date.

S90. Expanding roles of repeats

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 578. Characterization of full-length CNBP expanded alleles in myotonic dystrophy type 2 patients by Cas9-mediated enrichment and nanopore sequencing

Authors:

M. Delledonne[1], M. Alfano[1], L. De Antoni[1], F. Centofanti[2], V. V. Visconti[2], S. Maestri[1], C. Degli Esposti[3], R. Massa[2], M. D'Apice[2], G. Novelli[2], A. Botta[2], M. Rossato[1]; [1]Univ. Degli Studi Di Verona, Verona, Italy, [2]Univ. of Rome Torvergata, Rome, Italy, [3]Univ. of Verona, Verona, Italy

Abstract:

Myotonic dystrophy type 2 (DM2) is caused by CCTG repeat expansions in the *CNBP* gene, comprising 75 to >11,000 units and featuring extensive mosaicism, making it challenging to sequence fully-expanded alleles. To overcome these limitations, we used PCR-free Cas9-mediated nanopore sequencing to characterize *CNBP* repeat expansions at the single-nucleotide level in nine DM2 patients. The length of normal and expanded alleles can be assessed precisely using this strategy, agreeing with traditional methods, and revealing the degree of mosaicism. We also sequenced an entire ~50 kbp expansion, which has not been achieved previously for DM2 or any other repeat-expansion disorders. Our approach precisely counted the repeats and identified the repeat pattern for both short interrupted and uninterrupted alleles. Interestingly, in the expanded alleles, only two DM2 samples featured the expected pure CCTG repeat pattern, while the other seven presented also TCTG blocks at the 3′ end, which have not been reported before in DM2 patients, but confirmed hereby with orthogonal methods. The demonstrated approach simultaneously determines repeat length, structure/motif and the extent of somatic mosaicism, promising to improve the molecular diagnosis of DM2 and achieve more accurate genotype-phenotype correlations for the better stratification of DM2 patients in clinical trials.

S90. Expanding roles of repeats

Location: Conv Ctr/Concourse Hall E/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 579. Long-read sequencing reveal s*de novo* mutations in repetitive regions of the genome

Authors:

M. Noyes[1], W. T. Harvey[1], K. M. Munson[1], K. Hoekzema[1], M. C. Zody[2], E. E. Eichler[1,3]; [1]Univ. of Washington, Seattle, WA, [2]New York Genome Ctr., New York, NY, [3]Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA

Abstract:

De novo mutations (DNMs) are essential for understanding human evolution and disease and contribute to at least 30% of autism cases. Typically, DNM discovery leverages short-read Illumina data that cannot be unambiguously aligned to repetitive regions of the genome, which are consequently excluded. We hypothesize that by using long-read data, we can accurately identify DNMs in some of the most repetitive regions of the genome, increasing our estimates of the average number of DNMs per child and of the human mutation rate. In this study, we selected five quads, or families of two parents and two children, one of whom is affected with simplex autism. For each quad, we generated Illumina and PacBio HiFi sequence of all four members and Oxford Nanopore Technologies (ONT) sequence of the two children. In total, we identified *de novo* variation in ten children using their Illumina and HiFi sequences aligned to both GRCh38 and T2T-CHM13v1.1 references, which have a difference of more than 34 Mbp of accessible sequence between them. To maximize sensitivity, we used two Illumina-specific variant calling pipelines and two HiFi-specific pipelines. All DNM calls were validated by examining the underlying sequence data across the three sequencing platforms. Every true DNM call has support in HiFi, ONT, and Illumina reads for the child, and is absent in HiFi and Illumina reads for each parent. Across all ten children, we identified 920 *de novo* SNVs and small indels (<20bp) on the autosomes, an increase of 33% from previously published DNMs discovered in Illumina data aligned to GRCh38. On average, this callset indicates an autosome-wide substitution-rate of $1.45x10^{-8}$ substitutions/base/generation. In the more sensitive callset generated by HiFi data, we recovered potential functional variation, including two missense variants not seen by Illumina callers, although we were not able to find a putative causative variant for any of the children affected with autism. In regions with less than ten percent divergence, we were able to recover 148 DNMs, none of which have been reported in previous Illumina-based studies. We also identified DNMs in centromeres (n=76) and segmental duplications (n=48). In these regions, we find elevated mutation rates of more than $2.00x10^{-8}$ substitutions/base/generation, indicating that they may in fact be hypermutable when compared to unique regions of the genome. In summary, long-read sequencing and validation recovers 33% more *de novo* variation compared to short-read sequencing, including in repetitive sequences, where the mutation rate could be more than 40% higher than genome-wide.

S91. Genetic and functional underpinnings of epilepsy

Location: Conv Ctr/Petree C/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 582. Computational and functional analysis identifies AP3D1 as a potential candidate gene associated with epilepsy

Authors:

T. Nagai[1], E. Knapik[2], E. Gamazon[2,3]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Clare Hall, Univ. of Cambridge, Cambridge, United Kingdom

Abstract:

Epilepsy is a chronic neurological disease that affects millions of people worldwide. Current treatments for epilepsy are focused on controlling seizure activity through the use of antiepileptic drugs (AEDs). However, despite the dozens of AEDs that have been developed, a large proportion of epilepsy cases remain drug resistant. Genomic analysis has proven to be useful tool for identifying novel candidate genes beyond the ion channel inhibitors and GABA potentiators that make up most AED targets. In our study we apply PrediXcan, a transcriptome-wide association study (TWAS) method, on summary statistics from a previous epilepsy genome-wide association study (GWAS) done by the International League Against Epilepsy in order to identify genes associated with epilepsy based on predicted expression. The machine learning models used for this analysis were trained using expression quantitative trait loci (eQTL) data from the GTEx consortium combined with tissue similarity data from ENCODE and Roadmap consortiums. We then used quantitative polymerase chain reaction (qPCR) data in a zebrafish seizure model as a method to identify functionally relevant genes from the TWAS results. PrediXcan identified several genes associated with epilepsy, both known epilepsy genes such as GABRA2 ($p=2.67x10^{-7}$) and STX1B ($p=1.92x10^{-6}$), as well as novel genes such as CDK5RAP3 ($p=7.63x10^{-10}$), C1QL3 ($p=2.59x10^{-8}$), and AP3D1 ($p=4.06x10^{-7}$), all of which were significant after false discovery rate correction. The qPCR experiment then found that ap3d1 exhibited over two-fold decreased expression in seizure-induced zebrafish, indicating that AP3D1 expression is reduced in seizure response pathways. A transient ap3d1 knockout zebrafish model generated using CRISPR-Cas9 also exhibited increased sensitivity to the seizure inducing agent pentylenetetrazole (PTZ), providing further functional evidence that AP3D1 is involved in underlying seizure mechanisms. This combination of computational and functional data shows that AP3D1 is a promising candidate gene whose expression is directly involved with seizure activity.

S91. Genetic and functional underpinnings of epilepsy

Location: Conv Ctr/Petree C/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 583. Haploinsufficiency of *CHASERR*, a human long non-coding RNA, causes a severe neurodevelopmental disease, implicating dosage sensitivity of *CHD2* in brain development

Authors:

V. Ganesh[1,2,3,4], K. Riquin[5], K-M. Lamar[6], J. K. Goodrich[2], M. C. Aziz[6], K. V. Garimella[2], E. England[2], B. Weisburd[2], F. Aguet[2], C. A. Bacino[7], D. R. Murdock[7], H. Dai[7], J. A. Rosenfeld[7], L. T. Emrick[7], S. Ketkar[7], S. Kayani[8], B. Broadbent[9], B. Isidor[10], B. Cogné[10], D. MacArthur[11], I. Ulitsky[12], G. L. Carvill[6], A. O'Donnell-Luria[2,3,4]; [1]Dept. of Neurology, Brigham and Women's Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Div. of Genetics and Genomics, Dept. of Pediatrics, Boston Children's Hosp., Boston, MA, [4]Harvard Med. Sch., Boston, MA, [5]INSERM, CNRS, Univ.ire de Nantes, l'institut du thorax, Nantes, France, [6]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [7]Baylor Coll. of Med., Houston, TX, [8]Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [9]Coalition to Cure CHD2, Dallas, TX, [10]Ctr. Hosp.ier Univ.ire de Nantes, Nantes, France, [11]Ctr. for Population Genomics, Murdoch Children's Res. Inst., Melbourne, Australia, [12]Dept.s of Biological Regulation and Molecular NeuroSci., Weizmann Inst. of Sci., Rehovat, Israel

Abstract:

De novo loss-of-function variants in chromodomain helicase DNA-binding protein 2 (*CHD2*) are known to cause developmental and epileptic encephalopathy (Carvill et al., 2013). *CHASERR* is a conserved human long non-coding RNA (lncRNA) adjacent to *CHD2* on chromosome 15, and its mouse homolog (*Chaserr*) mediates *cis*-acting transcriptional repression of *Chd2* (Rom et al., 2019). Here we report two unrelated individuals with neurodevelopmental delay each harboring an ultra-rare heterozygous *de novo* deletion in the *CHASERR* locus. We report clinical similarities and distinctions between these two individuals compared with the known phenotypic spectrum of *CHD2* haploinsufficiency. We demonstrate reduced *CHASERR* mRNA expression in both individuals and corresponding overexpression of *CHD2* mRNA and protein in patient-derived cell lines, with allelic imbalance towards the CHD2 transcript in *cis* with the *CHASERR* deletion. We show for the first time that structural variants, including Alu-mediated deletions, that encompass non-coding elements (lncRNAs) are implicated in rare neurodevelopmental disorders through a pathogenic mechanism of altered *CHD2* gene dosage.

S91. Genetic and functional underpinnings of epilepsy

Location: Conv Ctr/Petree C/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 584. Missense variants in *RPH3A* cause defects in synaptic function and are associated with a neurodevelopmental disorder characterized by epilepsy

Authors:

L. Pavinato[1], J. Stanic[2], M. Barzasi[2], I. Eberini[2], M. Diluca[2], A. Gurgone[1], G. Chiantia[1], A. Marcantoni[1], E. Biamino[1], M. Spada[1], S. Hiatt[3], V. Cipriani[4], L. Vestito[4], D. Smedley[4], S. M. Sisodiya[5], Genomics England Research Consortium, A. Bruselles[6], S. De Rubeis[7], J. D. Buxbaum[7], M. Tartaglia[8], M. Giustetto[1], F. Gardoni[2], A. Brusco[1,9]; [1]Univ. of Turin, Turin, Italy, [2]Univ. of Milan, Milan, Italy, [3]HudsonAlpha, Huntsville, AL, [4]Queen Mary Univ. of London, London, United Kingdom, [5]UCL Queen Square Inst. of Neurology, London, United Kingdom, [6]Istituto Superiore di Sanità, Rome, Italy, [7]Icahn Sch. of Med. at Mount Sinai, New York, NY, [8]Ospedale Pediatrico Bambino Gesù, Rome, Italy, [9]Città della Salute e della Scienza Univ. Hosp., Turin, Italy

Abstract:

Exocytosis of neurotransmitters and synaptic retention of receptors are two key steps in synaptic neurotransmission. Rabphilin 3A (*RPH3A*; MIM *612159) is a small protein involved in the regulation of exo- and endocytosis processes at pre- and post-synapses and enriched at the level of the dendritic spines. RPH3A stabilizes synaptic N-methyl-D-aspartate (NMDA)-type glutamate receptors at cell surface through the formation of a ternary complex with PSD95 and the GluN2A subunit of the receptor, a fundamental step for the correct activation of the signal following the long-term potentiation induction, proper synaptic adaptation, synaptic plasticity, and cognitive behaviour. Through exome sequencing, we identified the *de novo* variant T450S in *RPH3A* in one individual affected by a neurodevelopmental disorder with untreatable epileptic seizures and the variant N618S in an individual showing high functioning autism spectrum disorder. By 100,000 Genomes Project we then identified three additional cases with the heterozygous missense variants E148D, R209K and Q508H, with a similar phenotype to the individual carrying T450S. A structural bioinformatics study of all the variants was performed, followed by a functional analysis of the impact of T450S and N618S variants, associated with seizures and autism respectively. Experiments on rat primary hippocampal neuronal cultures, showed that both the variants caused a reduced PSD95/GluN2A colocalization, and a significant increase in surface levels of GluN2A was observed for T450S. These results indicate that GluN2A-containing NMDARs are located mainly at the extra synaptic sites, suggesting that their stimulation can promote neurotoxicity events. Preliminary data, also showed a decrease in Calcium signal specifically at the level of synapsis. Interestingly, electrophysiological recordings showed increased GluN2A-dependent NMDAR currents for both variants. Neurons transfected with T450S also showed altered dendritic spines morphology namely decreased spine head width, with a significant increase of thin spines. Altogether, the presented experimental data showed functional impact for both the tested variants; however the impairment on neuronal function mediated by T450S was more pronounced, likely resembling the more severe phenotype of the patient. Overall, our data show that missense variants in *RPH3A* are involved in onset of neurodevelopmental disorders, and their impact on neuronal function can be various and associate with a different range of disease severity. We propose *RPH3A* as a novel candidate gene for a neurodevelopmental disorder associated with epilepsy.

S91. Genetic and functional underpinnings of epilepsy

Location: Conv Ctr/Petree C/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 585. *De novo* and biallelic variants in *R3HDM1*, encoding an Encore-like RNA binding protein hosting the microRNA MiR-128-1, disrupt cortical development and lead to neurodevelopmental phenotypes and epilepsy

Authors:

M. Scala[1,2], R. Ethiraj[3], C. Marini[3], D. Richter[3], M. Severino[2], M. Iacomino[2], S. Uccella[2], M. Mancardi[2], T. Feyma[4], J. Rosenfeld[5,6], A. Larson[7], R. Oegema[8], J. Saris[9], M. Ambrozkiewicz[10], V. Capra[2], F. Zara[2], P. Striano[2,1], A. Kaindl[3], G. Wulczyn[11]; [1]Univ. of Genoa, Genoa, Italy, [2]IRCCS Istituto Giannina Gaslini, Genoa, Italy, [3]Charité - Univ.smedizin, Berlin, Germany, [4]Gilette Children's Specialty Hlth.care, Saint Paul, MN, [5]Baylor Coll. of Med., Houston, TX, [6]Baylor Genetics Lab., Houston, TX, [7]Children's Hosp. Colorado, Univ. of Colorado Sch. of Med., Aurora, CO, [8]UMC Utrecht, Utrecht, Netherlands, [9]Erasmus Med. Ctr., ROTTERDAM, Netherlands, [10]Max Planck Inst. of Experimental Med., Gottingen, Germany, [11]Charité - Univ.smedizin Berlin, Berlin, Germany

Abstract:

Abstract
Background MicroRNAs (miRNAs) are endogenous, evolutionarily-conserved, non-coding RNAs with a crucial post-transcriptional regulatory role in the developing brain. MiR-128 modulates neuronal proliferation, differentiation, and excitability. One of the two isoforms, MiR-128-1, lies within intron 18 of the host gene *R3HDM1*, encoding a putative *Drosophila* Encore-like RNA-binding protein (RBP). The *R3hdm1* mRNA is upregulated during neuronal differentiation of embryonic cortical progenitors in the mouse, similar to its paralog *ARPP21* hosting the MiR-128-2 isoform[1-3].
Methods Using gene matching platforms, we assembled a cohort of four affected individuals from four unrelated families with a hitherto unreported neurodevelopmental disorder. Exome sequencing (ES) was performed to identify the underlying genetic cause. To investigate R3HDM1 function during corticogenesis, we compared the phenotypes of shRNA-mediated knockdown of both R3hdm1 and Arpp21 after *in utero* electroporation of upper layer (UL) progenitors. We also performed the inverse experiment and overexpressed R3hdm1.
Results All patients showed mild-to-profound neurodevelopmental impairment associated with heterogeneous epileptic phenotypes and brain abnormalities suggestive of malformations of cortical development (MCDs). ES led to the identification of six distinct variants in *R3HDM1*, either in biallelic or heterozygous status. We observed a reduction of successful migration into the upper layers (UL) after either R3hdm1 or Arpp21 knockdown, resembling the phenotype associated with miR-128 overexpression[4]. R3hdm1 overexpression led instead to enhanced entry and positioning at the top of the UL, matching the inhibition of miR-128 during migration[4]. Overmigration was associated with increased apical branching, suggesting functional redundance of R3HDM1 and ARPP21 in the promotion of dendritic growth[3].
Conclusion Our experiments highlight the functional crosstalk between R3HDM1, ARPP21 and miR-128 during cortical development in the mouse. These findings strongly suggest that pathways governing neuronal positioning and growth may be adversely affected in patients harboring variants in *R3HDM1*, underlying a novel neurodevelopmental disorder with epilepsy and intellectual disability.

S92. The past matters for cancer treatment and risk

Location: Conv Ctr/Room 502/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 588. Examining genetic susceptibility to anthracycline-related cardiomyopathy in cancer survivors using a gene-level approach

Authors:

N. Sharafeldin[1], L. Zhou[1], P. Singh[1], D. Crossman[1], X. Wang[2], L. Hageman[1], W. Landier[1], J. Blanco[3], P. Burridge[4], Y. Sapkota[5], Y. Yasui[5], G. Armstrong[5], L. Robison[5], M. Hudson[5], K. Oeffinger[6], E. Chow[7], S. Armenian[8], D. Weisdorf[9], S. Bhatia[1]; [1]Univ of Alabama at Birmingham, Birmingham, AL, [2]Univ. of North Texas, Denton, TX, [3]The State Univ. of New York at Buffalo, Buffalo, NY, [4]Northwestern, Chicago, IL, [5]St. Jude Children's Res. Hosp., Memphis, TN, [6]Duke Univ., Durham, NC, [7]Fred Hutchinson Cancer Ctr., Seattle, WA, [8]City of Hope, Durate, CA, [9]Univ. of Minnesota, Minneapolis, MN

Abstract:

The inter-individual variability in anthracycline-related cardiomyopathy risk among childhood cancer survivors has been attributed to associations with single nucleotide polymorphisms (SNPs). Considering effects of multiple SNPs on a gene and their interactions, however, remains unexamined. We used whole exome sequencing data to examine gene-level associations with cardiomyopathy among survivors of childhood cancer. For discovery, a matched case-control set of 278 childhood cancer survivors (129 cases; 149 controls) from COG-ALTE03N1 utilized logic regression to identify gene-level SNP combinations in 7,212 genes and ordinal logistic regression models to estimate gene-level associations with cardiomyopathy. Models were adjusted for primary cancer diagnosis, age at cancer diagnosis, sex, race/ethnicity, cumulative anthracycline dose, chest radiation, cardiovascular risk factors, and three principal components. Statistical significance threshold of 6.93×10^{-6} was used to account for multiple testing. Three independent cancer survivor populations were used to replicate gene-level and assess individual SNP associations: Childhood Cancer Survivor Study (CCSS) and Bone Marrow Survivors Study (BMTSS cohorts and a non-overlapping COG-ALTE03N1 case-control set. Median age at childhood cancer diagnosis for the discovery cases and controls was 6 and 8 years, respectively. Gene-level analysis identified statistically significant associations for *PR2X7* (OR=0.10; 95%CI: 0.04-0.27, $P=2.19 \times 10^{-6}$), *TNIK* (OR=4.58; 95%CI: 2.47-8.49, $P=1.34 \times 10^{-6}$), *LRRK2* (OR=0.19; 95%CI: 0.09-0.39, $P=6.62 \times 10^{-6}$), *MEFV* (OR=0.08; 95%CI: 0.03-0.24, $P=4.07 \times 10^{-6}$), *NOBOX* (OR=7.21; 95%CI: 3.23-16.1, $P=1.43 \times 10^{-6}$) and *FBN3* (OR=4.59; 95%CI: 2.42-8.71, $P=3.05 \times 10^{-6}$). The gene-level SNP combination on *P2RX7* was successfully replicated in the CCSS cohort (HR=0.65; 95%CI: 0.47-0.90, P=0.009). Individual SNPs across all significant genes, except *FBN3*, were associated with cardiomyopathy. *In silico* functional evidence supported the findings with biologically plausible links to inflammatory responses and cardiovascular disease in non-oncology populations. Specifically, P2RX7 protein antagonism has protective cardiovascular effects lowering blood pressure and atherosclerosis progression. Gene-level associations identified in this study have the potential to identify individuals at increased risk for cardiomyopathy and inform future discovery of therapeutic targets to mitigate this adverse outcome.

S92. The past matters for cancer treatment and risk

Location: Conv Ctr/Room 502/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 589. Cancer predisposition variants and subsequent-malignancy-related late-mortality among long-term survivors of childhood cancer

Authors:

C. Chen[1,2], N. Qin[3], M. Wang[2], S. Tithi[2], Y. Hui[2], W. Chen[2], G. Wu[2], D. Kennetz[2], M. N. Edmonson[2], M. C. Rusch[2], A. Thrasher[2], J. Easton[2], H. L. Mulder[2], N. Song[4,2], Q. Dong[2], K. Shelton[2], M. J. Ehrhardt[2], K. E. Nichols[2], W. Leisenring[5], Y. Yasui[2], S. Bhatia[6], G. T. Armstrong[2], K. K. Ness[2], M. M. Hudson[2], D. K. Srivastava[2], H. Wang[1], J. Zhang[2], L. L. Robison[2], Z. Wang[2]; [1]Shanghai Jiao Tong Univ., Shanghai, China, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Nanjing Med. Univ., Nanjing, China, [4]Chungbuk Natl. Univ., Cheongju, Korea, Republic of, [5]Fred Hutchinson Cancer Ctr., Seattle, WA, [6]The Univ. of Alabama System, Tuscaloosa, AL

Abstract:

BACKGROUND: We previously reported that among 3,006 survivors of childhood cancer in the St. Jude Lifetime Cohort (SJLIFE), 5.8% (95% CI = 5.0% to 6.7%) carried a cancer predisposition variant (CPV) in one of the 60 genes implicated in autosomal dominant cancer predisposition, and had increased risk of developing subsequent malignancies (SMNs). Among long-term survivors of childhood cancer, the rate of SMN-related late-mortality among carriers of CPV versus non-carriers is unknown. METHODS: Whole-genome sequencing (WGS) or whole-exome sequencing (WES) data for 12,475 five-year survivors were used to classify CPV status as previously described (Wang et al., JCO, 2018), including 4,402 from SJLIFE and 8,073 from the Childhood Cancer Survivor Study (CCSS; the CCSS original cohort data was downloaded from dbGaP, phs001327.v2). Mortality data were obtained through a search of the National Death Index. SMN-related mortality was analyzed using cmprsk R package implementing Fine & Gray method where other-cause mortality was treated as the competing risk. Deaths with unknown causes were excluded. RESULTS: A total of 263 SMN-related and 396 other-cause deaths occurred among the survivors. 642 (5.1%, 95% CI = 4.8% to 5.5%) survivors were CPV carriers. Overall, cumulative SMN-related mortality was significantly increased in CPV carriers versus non-carriers ($P = 2.9x10^{-7}$) whereas cumulative other-cause mortality was not (P = 0.38). By 40 years from achieving 5-year survival, the cumulative SMN-related mortality was 12.8% (95% CI = 10.0% to 15.6%) among CPV carriers as compared to 6.2% (95% CI = 5.7% to 6.7%) among non-carriers, and the cumulative other-cause mortalities were comparable (9.3% vs. 9.5%). After adjusting for genetically-determined race, sex, age at diagnosis and cancer treatment exposures, carrying a CPV was associated with increased rate of SMN-related mortality (Relative Rate [RR] = 3.34, 95% CI = 2.13 to 5.24, $P = 1.5x10^{-7}$ for the combined cohorts; RR = 5.24, 95% CI = 2.15 to 12.79, $P = 2.7x10^{-4}$ in SJLIFE; and RR = 3.02, 95% CI = 1.79 to 5.10, $P = 3.7x10^{-5}$ in CCSS). Notably, the association between CPV status and SMN-related mortality was substantially higher (RR = 4.82, 95% CI = 3.09 to 7.52) among those received higher doses of chest irradiation (no less than 20 Gy) than lower doses or none (smaller than 20 Gy) (RR = 1.45, 95% CI = 0.42 to 4.97), suggesting gene-treatment interactions. CONCLUSION: Carrying a CPV will not only increase the risk of developing SMN but also the risk of SMN-related mortality among long-term survivors of childhood cancer, highlighting the importance of genetic testing for CPVs to guide precision preventive survivorship care.

S92. The past matters for cancer treatment and risk

Location: Conv Ctr/Room 502/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 590. Genomic characterization of lymph node metastases in papillary thyroid carcinoma following the Chernobyl accident reveals an expression profile specific to the metastatic process

Authors:

O. Lee[1], D. M. Karyadi[2], C. Stewart[3], T. I. Bogdanova[4], S. Hartley[1], S. Schonfeld[1], M. Krznaric[5], M. Yeager[6], A. Hutchinson[7], B. Hicks[8], C. Dagnall[9], J. Gastier-Foster[10], J. Bowen[11], M. Machiela[1], E. Cahoon[1], K. Mabuchi[1], V. Drozdovitch[1], S. Masiuk[12], M. Chepurny[12], A. Berrington de González[1], G. Getz[3], G. A. Thomas[13], M. D. Tronko[4], L. M. Morton[1], S. J. Chanock[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Natl. Cancer Insititute, Rockville, MD, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]V.P.Komisarenko Inst. of Endocrinology and Metabolism of the Natl. Academy of Med. Sci. of Ukraine, Kyiv, Ukraine, [5]Charing Cross Hosp., London, United Kingdom, [6]FNLCR/NCI, Rockville, MD, [7]NCI Core Genotyping Facility, Bethesda, MD, [8]Leidos Biomed/DCEG/NCI/NIH, Bethesda, MD, [9]Natl. Cancer Inst. /Leido, Bethesda, MD, [10]Baylor Coll. of Med., Houston, TX, [11]Nationwide Children's Hosp., Columbus, OH, [12]Natl. Res. Ctr. for Radiation Med. of the Natl. Academy of Med. Sci. of Ukraine, Kyiv, Ukraine, [13]Imperial Coll. London, London, United Kingdom

Abstract:

Following the 1986 Chernobyl nuclear power plant accident, increased childhood exposure to radioactive iodine (^{131}I) in the surrounding regions, primarily through consumption of contaminated food sources, has been consistently associated with increased risk of developing papillary thyroid carcinoma (PTC). Higher frequency of cervical lymph node metastases (LNM) is well recognized in pediatric PTC, including pediatric cases following the Chernobyl accident, but the metastatic process is poorly understood.

We recently conducted a genomic landscape analysis of 440 cases of PTC that provided insights into radiation-associated molecular characteristics of PTC occurring after the accident (Morton et al., *Science* 2021). Here, we expand on that study with additional detailed clinical data to identify molecular and clinical predictors of LNM occurrence. We then conducted a comprehensive genomic landscape analysis of 47 LNMs with patient-matched primary PTCs to investigate the specific genomic alterations occurring in LNM.

Among the 440 cases, PTC with fusion drivers had substantially higher rates of LNM than PTC with mutation drivers (56% versus 30%, $P=2.2\times10^{-6}$), whereas $P>1.0\times10^{-4}$ for all other patient, clinical, and molecular characteristics, including radiation dose (P=0.31). We further observed striking heterogeneity by driver gene ($P=4.2\times10^{-20}$), with the highest rate of LNM occurrence among PTC with *RET* (73%) and other *RTK* (64%) fusion drivers but much lower LNM rates among PTC with fusion drivers in other genes (10%) or mutations in *BRAF* (37%), *RAS* (7%), or other genes (7%).

Molecular profiling of the 47 LNM revealed 100% concordance of the driver alterations and no novel drivers compared with the matched primary PTCs, as well as highly concordant mutational spectra. Transcriptome analysis revealed 17 differentially expressed genes in LNM compared to primary PTC ($P<1.0\times10^{-4}$). Most notably, we observed overexpression in LNM of *HOXC10* ($P=6.4\times10^{-23}$) as well as 8 other genes in the 12q13.13 *HOXC* locus, which has been linked to regulation of cell proliferation, tumor invasion, and promotion of TGF-beta signaling. The second strongest association was for reduced expression in LNM of *BRINP3* ($P=1.3\times10^{-17}$) as well as two other negative regulators within the TGF-beta pathway.

Our findings provide insights into the molecular processes underlying the development of metastatic PTC and underscore the critical role of the driver mutation in PTC metastasis. To better understand the biological underpinnings of metastatic PTC, further investigation of the altered expression of *HOXC* locus and its role in disrupting activity of TGF-beta is indicated.

S92. The past matters for cancer treatment and risk

Location: Conv Ctr/Room 502/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 591. Selection acting on somatic structural variation in blood impacts molecular function and cancer risk among humans

Authors:

K. Skead[1,2], M-J. Fave[1], D. Soave[3], V. Bruat[1], Q. Morris[4], P. Awadalla[1,5]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]Wilfrid Laurier Univ., Waterloo, ON, Canada, [4]Mem. Sloan Kettering Cancer Ctr., New York, NY, [5]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada

Abstract:

The age-associated accumulation of somatic mutations in blood, termed clonal hematopoiesis of indeterminate potential (CHIP), has been implicated in the development of blood cancers and cardiovascular conditions. Their impact on phenotypes have largely been captured as genetic epidemiological associations with a number of retrospectively captured traits in cohort studies. Here, we evaluate how natural selection shapes the prevalence of clonal somatic structural variants (SSVs) in blood, and their impact on cancers, among 15,910 individuals. Among participants of the Canadian Partnership for Tomorrow's Health and the Thousand Genomes Project, we find that SSV-inferred CHIP is three times as high as previously reported, with one in eight individuals in the population harboring a SSV in their blood. Using a novel statistical approach, we identify and characterize SSV hotspots across the genome, and find that SSVs are enriched for genes implicated in blood cancer development. In particular, we capture an enrichment of loss events at known haploinsufficient tumour suppressor genes suggesting that SSVs might be early arising cancer driver events. Correspondingly, we find that individuals who harbor at least one SSV are 3.5x more likely to progress to blood cancer and 1.8x more likely to progress to breast, prostate or pancreatic cancer. To capture the impact of SSVs on molecular phenotypes, we investigate the relationship between SSVs and the transcriptome. We show that gains, losses, and copy number (CN) neutral variants impact gene expression distinctly, with stabilizing selection shaping the penetrance of SSVs in gene expression. We perform SSV-adjusted eQTL mapping to reveal that one in three eQTLs are confounded by SSVs which impacts our understanding of how germline genetic factors contribute to expression variation, regulatory mechanisms and cancer outcomes. Our work shows how different classes of selection shape clonal dynamics in healthy and pre-cancerous blood thus enabling us to better understand why certain individuals are at risk of malignancy.

S93. Understanding GWAS signals: From variants to function

Location: Conv Ctr/Room 515/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 594. Cross-cohort eQTL fine-mapping utilizing TOPMed whole genome sequencing identifies tens of thousands of independent eQTLs signals and thousands of eQTLs colocalizing with complex trait-associated variants

Authors:

P. Orchard[1], F. Aguet[2], T. Blackwell[1], K. Ardlie[3], P. J. Castaldi[4], A. V. Smith[1], R. Joehanes[5,6], A. Saferali[7], H. E. Wheeler[8], C-T. Liu[9], M. Cho[7], C. Hersh[7], L. Mestroni[10], L. Kachuri[11], A. P. Reiner[12], X. Li[13], M. Taylor[10], D. A. Meyers[13], S. S. Rich[14], G. Abecasis[1], N. Heard-Costa[15], L. J. Scott[1], J. I. Rotter[16], H. Tang[17], D. Levy[5,6], L. M. Raffield[18], S. C. J. Parker[1], NHLBI TOPMed Consortium; [1]Univ. of Michigan, Ann Arbor, MI, [2]Illumina, Inc., San Diego, CA, [3]Broad Inst., Cambridge, MA, [4]Brigham and Women's Hosp., Cambridge, MA, [5]Natl. Heart, Lung, and Blood Inst., NIH, Bethesda, MD, [6]Framingham Heart Study, Framingham, MA, [7]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [8]Loyola Univ. Chicago, Chicago, IL, [9]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [10]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [11]Stanford Univ. Sch. of Med., Stanford, CA, [12]Univ. of Washington, Seattle, WA, [13]Univ. of Arizona, Tucson, AZ, [14]Univ. of Virginia, Charlottesville, VA, [15]Boston Univ. Sch. of Med., Boston, MA, [16]The Lundquist Inst. for BioMed. Innovation at Harbor-UCLA Med. Ctr., Torrance, CA, [17]Stanford Univ., Stanford, CA, [18]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract:

Most genetic variants associated with complex traits and diseases occur in non-coding genomic regions and are hypothesized to regulate gene expression. To understand the genetics underlying gene expression variability, we performed *cis* expression quantitative trait locus (*cis*-eQTL) analyses using RNA-seq and whole genome sequencing (WGS) data from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program from 6,602 whole blood samples of European (EUR; 68%), African (21%) and Indigenous American (10%) ancestry. Notably, this exceeds the sample size of published RNA-seq and WGS-based *cis*-eQTL analyses, which enabled us to test variants with minor allele frequency (MAF) below 0.01 and detect secondary signals for 15,317 genes.

At a MAF≥0.001, we identified 19,381 genes with at least one eQTL (5% FDR, testing variants within 1Mb of the transcription start site; 22,180 genes tested). We fine-mapped independent eQTL signals using the SuSiE method and identified 77,398 eQTL signals (95% credible sets; median 17,183 variants tested per gene and 3 credible sets discovered per gene), including 31,810 credible sets containing a single variant. By contrast, restricting to variants with higher MAF (MAF≥0.01), we identified 70,943 eQTL signals (median 7,953 variants tested per gene and 3 credible sets discovered per gene), and 29,690 95% credible sets containing a single variant.

To assess the utility of this dataset to identify target genes and nominate causal variants for genome wide association study (GWAS) signals, we colocalized independent *cis*-eQTL signals with 33,141 fine-mapped EUR GWAS signals from 172 UK Biobank traits. 5,782 GWAS signals colocalized with an eQTL (SuSiE-coloc PP4 posterior probability of colocalization > 0.8). Of these, 1,648 GWAS signals colocalized with an eQTL from more than one gene. Of 4,134 GWAS signals colocalizing with only one gene, in 52% of cases the gene was not the nearest gene. 2,910 of the 5,782 colocalizing GWAS loci colocalized with only secondary eQTL signals. We identified 215 instances in which multiple neighboring GWAS signals for a given trait colocalized with multiple eQTLs from the same gene. For example, in one 843kb window we identified six independent GWAS signals for neutrophil percentage, three of which are in or near *ACKR1* (previously shown to regulate neutrophil counts) and colocalize with three independent *ACKR1* eQTL signals (each with a single variant 95% eQTL credible set).

In summary, this dataset demonstrates the utility of large-scale WGS-based eQTL studies to map genetic regulatory effects on gene expression at unprecedented resolution and nominate causal genes for thousands of GWAS signals.

S93. Understanding GWAS signals: From variants to function

Location: Conv Ctr/Room 515/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 595. Integrative analysis of metabolite GWAS illuminates the molecular basis of pleiotropy and genetic correlation

Authors:

C. Smith[1], N. Sinnott-Armstrong[1], E. Fauman[2], J. Pritchard[1]; [1]Stanford Univ., Stanford, CA, [2]Pfizer, Cambridge, MA

Abstract:

Pleiotropy and genetic correlation are widespread features in genome-wide association studies (GWAS), but they are often difficult to interpret at the molecular level. Investigation of molecular traits in related pathways, with well-documented biology jointly impacting their levels, provides an opportunity to uncover putative molecular mechanisms. Here, we perform GWAS of 16 metabolites (n=94,464 individuals) clustered at the intersection of amino acid catabolism, glycolysis, and ketone metabolism in a subset of UK Biobank. We utilize the well-documented biochemistry jointly impacting these metabolites to analyze pleiotropic effects in the context of their pathways.

Among the 213 lead GWAS hits, we find strong enrichment for genes encoding pathway-relevant enzymes (n=68 variants, 25-fold enrichment, Poisson rate test P < 2e-16) and transporters (n=46 variants, 5.2-fold enrichment, P=9e-16). We show specific examples of candidate molecular mechanisms explaining the association of variants with multiple biologically-related metabolites. These include associations at *PDPR*, *SLC36A2*, and *PCCB*, where we show that the direction and magnitude of their effects are consistent with metabolite biochemistry and disease etiology.

We also find that the effect direction and local genetic correlation of variants affecting biology between metabolite pairs often contrast with those of upstream or downstream variants as well as the polygenic background. These "discordant" variants are more likely to affect enzymes and transporters than other gene types (4.1-fold, P=0.034).

Using a model derived from these results, we demonstrate that genome-wide correlations often hide striking variation in patterns of pleiotropy, often masking effects from variants and pathways most relevant to the core biology of the traits. For example, we find the local genetic correlation in pathways containing and adjacent to alanine and glutamine range from -0.77 to 0.67, despite an overall genetic correlation of 0.16. Finally, we leverage the interpretability of our metabolite GWAS results to evaluate the molecular basis of a noncoding coronary artery disease GWAS hit.

We use local genetic correlation methods to pinpoint overall heritability of cardiometabolic diseases within pathways to identify putative causal mechanisms. Together, this underscores the potential of unifying biochemistry with dense metabolomics data to understand the molecular basis of pleiotropy in complex traits.

S93. Understanding GWAS signals: From variants to function

Location: Conv Ctr/Room 515/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 596. Massively parallel reporter assays of QT interval GWAS enhancer variants

Authors:

A. Kapoor[1], D. Lee[2], A. Chakravarti[3]; [1]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Boston Children's Hosp., Boston, MA, [3]NYU Sch. of Med., New York, NY

Abstract:

Sequence variation in multiple *cis*-regulatory elements (CREs) modulating a target gene's expression is the leading hypothesis for GWAS signals in the noncoding genome. However, identification of such causal *cis*-regulatory variants requires experimental data. Here, we performed a massively parallel reporter assay (MPRA)-based screen to evaluate *cis*-regulatory activities of 729 biallelic variants that are common (1000 Genomes EUR MAF>1%), in high linkage disequilibrium with any of the independent hits (r^2>0.9) at QT interval GWAS loci, and overlap human cardiac open chromatin extended regions. Two pools of 200-base long oligos with a 20-base 5′ flank, variant-centered 129-base sequence, 18-base spacer with EcoRI and SbfI sites, 13-base unique barcode and a 20-base 3′ flank were designed. Barcodes were designed to contain all four nucleotides and GC-content of 0.4-0.6, and further avoid microRNA seed binding sites, ≥3-base homopolymers, the two restriction sites, and to have a 2 base minimum edit distance between any two barcodes. Each allele/variant was linked to 50 different barcodes. Oligo pools were cloned into KpnI-XbaI digested pGL4.23 to generate pre-reporter libraries, followed by cloning of a minimal promoter-driven eGFP cassette into the EcoRI-SbfI site of the pre-reporter libraries. We generated the final reporter plasmid libraries from >5 million clones each to preserve library complexity. The two libraries were transfected into the mouse cardiomyocyte cell line HL1 in 10 parallel replicates, each with 4 million cells. 48h post-transfection, cells were harvested to generate cDNA-based indexed libraries (10 replicates/pool) along with input plasmid-based indexed libraries (3 replicates/pool), followed by barcode sequencing. Read counts from the input plasmid libraries were aggregated across replicates. Barcode-level reporter activity in each replicate was measured as a $\log_2$-transformed ratio of sequence depth-normalized cDNA/input plasmid read counts, and averaged across replicates and applicable barcodes to measure test element-level activity. In a preliminary analysis, using a >30% increase in cDNA reads count over input-plasmid (for either reference or alternate test element) to define enhancers, 69 enhancer CREs were identified of which 45 showed significant allelic difference by *t*-test with multiple testing corrections using Benjamini-Hochberg procedure (FDR<0.01). These enhancer CRE variants map to 16 GWAS loci marked by genes including *NOS1AP*, *KCNQ1*, *KCNH2*, *SCN5A*, *PLN*, and *ATP1B1*. Our MPRA screen identified functional CRE variants are being evaluated for causality by explaining target gene expression variation.

S93. Understanding GWAS signals: From variants to function

Location: Conv Ctr/Room 515/West Building

Session Time: Saturday, October 29, 2022, 10:30 am - 11:30 am

ProgNbr 597. From variants to functions for coronary artery disease: Systematic Perturb-seq links GWAS loci to disease programs in endothelial cells

Authors:

J. Engreitz[1], G. R. Schnitzler[2], H. Kang[1], R. Ma[1], V. Lee-Kim[3], A. Zeng[1], S. Vellarikkal[3], R. Zhou[1], O. Sias-Garcia[3], G. Munson[2], K. Guo[1], B. Cleary[2], E. S. Lander[2], R. M. Gupta[3]; [1]Stanford Univ., Stanford, CA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Brigham and Women's Hosp., Boston, MA

Abstract:

Genome-wide association studies (GWAS) have discovered >200 associations for coronary artery disease (CAD), each of which could point to genes and pathways that influence disease risk. It is thought that a fraction of these CAD risk loci influences the functions of endothelial cells, and that genes in multiple GWAS loci might act together in certain pathways. Yet, identifying these genes and pathways has proven challenging: each GWAS locus can have 2-20 candidate genes, a gene may participate in one or more pathways in a given cell type, and it remains unclear which genes and pathways would be likely to influence disease risk. Here, we sought to create an unbiased catalog of gene pathways and their regulators in endothelial cells to link CAD risk variants to functions. To do so, we applied CRISPRi-Perturb-seq to knock down the expression of all genes within 500 Kb of all coronary artery disease GWAS loci (2,300 genes in total) and measure their effects on the transcriptome using single-cell RNA-seq. We used consensus non-negative matrix factorization to define 60 gene expression programs—including core cellular programs such as ribosome biogenesis and endothelial cell-specific programs such as flow response and angiogenesis—and link these programs to upstream regulators including transcription factors, chromatin regulators, metabolic enzymes, and signaling cascades. By combining this gene-to-program catalog with variant-to-gene enhancer maps, we find that candidate CAD genes converge onto 6 interrelated gene programs, together involving known and novel genes in 35 of 229 CAD GWAS loci. Analysis of these programs revealed that the cerebral cavernous malformations (CCM) complex—whose potential connection to coronary artery disease has not been previously explored—acts upstream to regulate other CAD genes involved in cytoskeletal organization, extracellular matrix remodeling, and cell migration. The strongest regulator of these programs is a highly conserved but poorly studied gene that we show acts together with the CCM complex. Together, our study nominates new genes that likely influence risk for CAD, identifies convergence of CAD risk loci into certain gene programs in endothelial cells, and demonstrates a generalizable strategy to catalog gene programs to connect disease variants to functions.